

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 23, 2003, 11:24:35 ; Search time 2154 seconds
(without alignments)
8079.609 Million cell updates/sec

Title: US-09-920-953-2

Perfect score: 598
Sequence: 1 gcgtgcctgcagacgcggc.....agccatttcgcaccagcc 598

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

total number of hits satisfying chosen parameters:	4109280
--	---------

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:★

1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sv:*
13: gb_un:*
14: gb_vt:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_ov:*
22: em_or:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vt:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rnd:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_hgo_hum:*
40: em_hgo_mus:*
41: em_hgo_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query			DB	ID	Description
		Match	Length				
1	81.4	13.6	1372	8	CEL1410	X72910 C.eugametos	
2	79	13.2	587	3	TETHEMOCP	X72920 Tetrahymena	
3	77.6	13.0	1197	8	CEL1637P	X72916 C. eugametos	
4	76.6	12.8	481	1	AF475938	AF475938 Synechocyst	
5	70.4	11.8	146174	1	D90910	D90910 Synechocyst	
6	53.8	9.0	494	3	TETHEMOCT	X72919 Tetrahymena	
7	50.8	8.5	3605	1	MKDNATHS	Z50745 M.kandleri	
8	50.8	8.5	12792	1	AE010389	AE010389 Mouse delta	
9	48.2	8.1	1898	10	MUSDELTA	M74590 Mouse delta	
10	48.2	8.1	2330	10	MUSCRSP	M73963 Mus musculus	
11	48.2	8.1	3041	10	MUSSTRANS01	LL1969 Mouse delta	
12	48	8.0	234817	10	AL663048	AL663048 Mouse DNA	
C 13	47.6	8.0	110000	2	LMFLCHR32_06	Continuation (7 of	
C 14	47.6	8.0	110000	2	LMFLCHR36_07	Continuation (8 of	
C 15	47.4	7.9	137936	2	CNSORCAX	AL844874 Oryza sat	
C 16	47.4	7.9	144778	2	CNSOTFEF	AL513404 Oryza sat	
17	47.2	7.9	3014	3	AY084205	AY084205 Drosophil	
18	47	7.9	10486	1	BJDNAKJ1	Y09633 Bradyrhizob	
19	46.4	7.8	411	1	MSM249336	AJ249386 Mycobacte	
20	46.4	7.8	1737	1	AF130980	AF130980 Mycobacte	
C 21	46.4	7.8	16563	1	AE007025	AE007025 Mycobacte	
C 22	46.4	7.8	35377	1	MTCY48	Z74020 Mycobacteri	
C 23	46.4	7.7	117080	2	AC098840	AC098840 Magnapor	
24	45.8	7.7	135216	8	AC069145	AC069145 Oryza sat	
C 25	45.8	7.7	144724	8	AC078948	AC078948 Oryza sat	
C 26	45.2	7.6	149117	2	AC120527	AC120527 Oryza sat	
27	45	7.5	11401	1	D50308	D50308 Lysobacter	
28	44.8	7.5	5520	3	AC084329	AC084329 Leishmani	
29	44.6	7.5	11034	1	AE005065	AE005065 Halobacte	
30	44.4	7.4	154494	2	AP005070	AP005070 Oryza sat	
C 31	44.2	7.4	11509	1	AE012381	AE012381 Xanthomon	
32	44	7.4	12172	1	AE012305	AE012305 Xanthomon	
33	44	7.4	79390	2	AC020442	AC020442 Drosophil	
C 34	44	7.4	84653	3	AC004290	AC004290 Drosophil	
35	44	7.4	187350	2	AC129717	AC129717 Oryza sat	
36	44	7.4	180814	3	AC099023	AC099023 Drosophil	
C 37	44	7.4	256764	3	AE003792	AE003792 Drosophil	
C 38	43.8	7.3	32274	1	SCC8A	AF136892 Streptomy	
39	43.4	7.3	867	1	AF145230	AF145230 Ralstonia	
C 40	43.4	7.3	110000	2	LMFLCHR36_03	Continuation (4 of	
41	43.4	7.3	17466	1	AE001826	AE001826 Deinococc	
C 42	43.2	7.2	13461	14	AF084533	AF084543 Tupalia he	
C 43	43.2	7.2	195859	14	AF281817	AF281817 Tupalia he	
C 44	43	7.2	20127	1	AC124194	AC124194 Mus muscu	
C 45	43	7.2	213050	1	AL646067	AL646067 Ralstonia	

ALIGNMENTS

RESULT 1	1372 bp	mrna	linear	PLN 16-JUN-1994
CEL1410				
LOCUS				
DEFINITION				
ACCESSION				
VERSION				
KEYWORDS				
SOURCE				
ORGANISM				
REFERENCE				
AUTHORS				
TITLE				

green alga *Chlamydomonas eugametos*
 Mol. Gen. Genet. 243 (2), 185-197 (1994)
 94232186
 8177215
 REFERENCE
 2 (bases 1 to 1372)
 AUTHORS
 Guertin, M.
 TITLE
 Direct Submission
 Submitted (23-MAR-1993) M. Guertin, Laval University, Department of
 Biochemistry, Pavillion Vachon Room 3426, Quebec, G1K 7P4, CANADA
 JOURNAL
 Biochemistry, Pavillion Vachon Room 3426, Quebec, G1K 7P4, CANADA
 FEATURES
 source
 1. .1372
 Location/Qualifiers
 /organism="Chlamydomonas eugametos"
 /strain="UTEX9"
 /db_xref="taxon:3053"
 /clone="lambda 21.1.1"
 /clone_lib="cDNA library in lambda ZAP"
 84. .599
 gene
 84. .599
 CDS
 84. .599
 /gene="Li410"
 /codon_start=1
 /product="Li410p"
 /protein_id="CAA51420.1"
 /db_xref="GI:437981"
 /translation="NMRTVQLRLRPCIRAOQPVRAPTSVAAPATTPAPTKKCPFS
 LFAKLGREAVEAAVDKFNKVAADPTVSFVSKDMKVRQSKQPAFLAYALGGAEEW
 RGDNRTHAKDLVPHLTDVHFQAVVRHLSDTLAELGVTPGDIADAMAVVASTKTEVLN
 MPROGAGSNR"
 BASE COUNT 278 a 421 c 377 g 296 t
 ORIGIN
 Query Match 13.6%; Score 81.4; DB 8; Length 1372;
 Best Local Similarity 52.1%; Pred. No. 2.5e-06;
 Matches 209; Conservative 0; Mismatches 186; Indels 6; Gaps 1;
 QY 1 GGGCTGCTGCGAGCGCGCCACACAGACAGACGCGGAGCGGATGCGGCGC 60
 DB 150 GCGCCTACGTCGCTGCGCCCTGCCACTGCCACACACACCTGCGCCACTAAGAGTGC 209
 QY 61 AAGAAGCTCTTTGATGACCTGGCGCGCGAGAGGCGATCAAGCTGCGGTTGACACCTTC 120
 DB 210 TTCTCNCCTCTTCCGACGTCGGCGCGCGAGAGGCGTGAAGAGCGGTCGACAGTTT 269
 QY 121 TAGGATAAGGTGCTGCTGACCGCGAGCTGCTGCCCTTCTCGAGTCCCTGGACATGCAA 180
 DB 270 TACAACAAGTTTGTGCGCCACCCACCGTCAGCGTCTTCTTCTCCAAGAGCGACATGA 329
 181 GAGCAGAGATGAAGCAGTCAAGTTCATGAGCTTCGTTTGGCGGAGCAGACCAATAC 240
 DB 330 GTGCAGCGCTCCAAGCAGTTTCGTTCTCTGGCTATGCTCTGGTGGTGCAGCCGAGTGG 389
 QY 241 AAGGCGCGAAGCATGTACACGACACAGCGCCCATCTGGTCAAGGGCCACGCGCTGGACCA 300
 DB 390 AAGGCGAGGACATGCGGACCGACACACAGGACCTGGT-----CCCTCACCTCACCGAC 443
 QY 301 CCCACTTTGCAAGATCAAGCAGTACCTTTGGAGAGACGCTGCAAGAGATGGCGCTCAAG 360
 DB 444 GTCCACTTCCAGCGCGTAGTCCGTCACCTGAGCGACACCTTGTGAGCTGGCGCTCACC 503
 QY 361 CAGGATGTGATCCAGCAGCGCGCGGAGTGGTGGAGTCCAC 401
 DB 504 CTGGCGCATCGCTGATGCCATGCGCAGTGGTGGCGCTCTAC 544

RESULT 2
 TETHMOGP
 LOCUS
 TETHMOGP 587 bp mRNA linear INV 27-OCT-2000
 DEFINITION
 Tetrahymena pyriformis mRNA for hemoglobin.
 ACCESSION
 D13920
 VERSION
 D13920.1 GI:217409
 KEYWORDS
 hemoglobin.
 SOURCE
 Tetrahymena pyriformis cDNA to mRNA.

ORGANISM Tetrahymena pyriformis
 Eukaryota; Alveolata; Ciliophora; Oligohymenophorea;
 Hymenostomatida; Tetrahymenina; Tetrahymena.
 REFERENCE
 1 (bases 1 to 587)
 AUTHORS
 Takagi,T., Iwaasa,H., Yuasa,H., Shikama,K., Takemasa,T. and
 Watanabe,Y.
 TITLE
 Primary structure of Tetrahymena hemoglobins
 JOURNAL
 Biochim. Biophys. Acta 1173 (1), 75-78 (1993)
 MEDLINE
 93250050
 REFERENCE
 2 (bases 1 to 587)
 AUTHORS
 Takagi,T.
 TITLE
 Direct Submission
 JOURNAL
 Submitted (09-DEC-1992) Takashi Takagi, Biological Institute,
 Faculty of Science, Tohoku University; Aobayama, Sendai, Miyagi
 980, Japan (E-mail:TAKAGI@PNTOHOK.BITNET,
 Tel.:022-222-1800(ex.3469), Fax:022-263-9206)
 FEATURES
 source
 1. .587
 Location/Qualifiers
 /organism="Tetrahymena pyriformis"
 /db_xref="taxon:5908"
 17. .382
 CDS
 /codon_start=1
 /trans_cable=6
 /product="hemoglobin"
 /protein_id="BAA03015.1"
 /db_xref="GI:418106"
 /translation="MNKPTIVKLGGENAKAAVPLEYKVLADERYVHFKNFTDMD
 HOTQQQDFLMLLGGPNHYGKNTEAHKGNLQNLHFDLIENLAATLKLGLGTVDA
 VINEAKVIEHTRKMDLCK"
 polyA_signal 442. .447
 polyA_signal 482. .487
 BASE COUNT 198 a 121 c 104 g 164 t
 ORIGIN
 Query Match 13.2%; Score 79; DB 3; Length 587;
 Best Local Similarity 54.1%; Pred. No. 8.2e-06;
 Matches 191; Conservative 0; Mismatches 150; Indels 12; Gaps 1;
 QY 58 CCACAAGAAGCTCTTTGATGACCTGGCGCGCAGAGGCGATCAAGCTGCGGTTGCACAC 117
 DB 26 CCCCAACTATTTATGAAGAAGCTCGAGCGGAAATGCCATGAAGCTCGCTCCCTC 85
 QY 118 TTCTACGATAAGGTGCTGGCTGACCGGAGCTGCTGCCCTTCTTGAGTCCCTGGACATG 177
 DB 86 TTCTACAAGAAGTCTTAGCTCATGAAGAGTCAAGCATTTCTTCAAGAACACCGCATG 145
 QY 178 CAAGAGCAGAGATCAAGCAGGTCAAGTTCATGAGCTTCGTTTGGCGGAGCAGACCAA 237
 DB 146 GATCACCAAAACCAAGCAATAAACTGACTTCTCTCACCATGCTCTTAGTGGTCCCAACCAT 205
 QY 238 TACAAGGCGGAGGATGTACGACGACACAGCGCCCATCTGGTCAAGGGCCACGCGCTGGAC 297
 DB 206 TACAAGGTTAAATAATGACTGAAGCTCA-----CAAGGTTAAGCTTGGCA 253
 QY 298 CACCGCCACTTTGCAAGATCAAGCAGTACCTTTGGAGAGACGCTGCAAGAGATGGCGTC 357
 DB 254 AACTTGCATTTGATGCCATCATTCGAAACCTTGTGCTACCTTAAAGAGGCTCGTGC 313
 QY 358 AAGCAGGATGTCATCCAGCAGCGCGCGGAGTGGTGGAGTCCACCGCGGACGA 410
 DB 314 ACCGATGCTGTATTAAAGAGGCTCTAAGGTCTATCGAACACACCGCGTAAGA 366

RESULT 3
 CELI637P
 LOCUS
 CELI637P 1197 bp mRNA linear PLN 16-JUN-1994
 DEFINITION
 C.eugametos Li637p mRNA.
 ACCESSION
 X72916
 VERSION
 X72916.1 GI:437982
 KEYWORDS
 haemoglobin; Li637 gene; light-induced expression; nuclear gene.
 SOURCE
 Chlamydomonas eugametos.
 ORGANISM
 Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;

REFERENCE AUTHORS

TITL, E

JOURNAL
MEDLINE

REFERENCE

TITLE:

JOURNAL
MEDLINE

REFERENCE

TITLE
JOURNAL

COMMENT

FEATURES

gene

CDs

gene

CDS

gene

50

1 Kaneko, T., Tanaka, A., Sato, S., Kotani, H., Sazuka, T., Miyajima, N., Sugura, M. and Tabata, S. Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp. strain PCC6803. I. Sequence features in the 1 Mb region from map positions 6% to 9% of the genome. DNA Res. 2 (4), 153-166 (1995) 9613759

2 Kaneko, T., Sato, S., Kotani, H., Tanaka, A., Asamizu, E., Nakamura, Y., Miyajima, N., Hirose, M., Sugiyama, M., Sasamoto, S., Kimura, T., Hosouchi, T., Matsuno, A., Muraki, A., Nakazaki, N., Naruo, K., Okumura, S., Shimpo, S., Takeuchi, C., Wada, T., Watanabe, A., Yamada, M., Yasuda, M. and Tabata, S.
Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp. strain PC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions
DNA Res. 3 (3), 109-136 (1996)

3 (bases 1 to 146174)
Tabata S.
Direct Submission
Submitted (28-JUN-1996) Satoshi Tabata, Kazusa DNA Research
Institute, The First Laboratory for Plant Gene Research, Yana
1532-3, Kisarazu, Chiba 292-0812, Japan
(E-mail: tabata@kazusa.or.jp, URL: <http://www.kazusa.or.jp/cyano/>,
Tel: +81-438-52-3933 (ex.2330), Fax: +81-438-52-3934)
Potential protein coding regions were assigned on the basis of
similarity search of the ORFs and GeneMark analysis.

```

Location/Qualifiers
1. .146174
/organism="Synecocystis sp. PCC 6803"
/strain="PCC6803"
/db_xref="taxon:1148"
/note="synonym:Synecocystis PCC6803"
1. .486
/gene="apca"
1. .486
/gene="apca"
/note="ORF_ID:slr2067"
/codon_start=1
/transl_table=11
/product="allophycocyanin a chain"
/protein_id="BA17874.1"
/db_xref="GI:1652957"
/translation="MSIVTKSLVNDAEAERYLSPGELDRIKAFVTVGGAARLRIAETLT
GSRTIVKQAGDRFLQKRPDIVSPGNGAYGEMTATCLRDMYYLLRLVTVYVGVSGDV
PIEILGVVREMYRISLGTPIEAVQAQSVREMKVEASGLMSSDDAAASVYFVYVIGKM

```

/translation="MRMFRTACVPSQTRIQTQRELQNTYFTTKLVPYDNMFQQRIRI
KMGGRIVKVELATRGPGTNAGLA"
complement(1578..2480)
/gene="prma"
complement(1578..2480)
/gene="prma"
/note="ORF_ID:sll1909"
/codon_start=1
/transl_table=11
/product="ribosomal protein l11 methyltransferase; PrmA"
/protein_id="BAAL1877.1"
/db_xref="GI:1652960"
/translation="MAAASMWEIFRILCHPSLEETAFWRLKFGCIGCTSTPEKKAHSL
VGRYLPQEKAEILDLAALACEQDALLQVQPKPRFHQWILIDEDNASISWKEHWPOTPP
VGDRIIFPAWIDPPENSRLILRLDPGVAQGTTHATQCLLESLEMRVEPKHCVL
LDLQCGSGILGIGVALGAARKAYGVGNDPLTVESARHNRLHQIHFDNLVINGSGVPEP
LEQLIAEPVDRGICNLIAEVIYVDLLPFLPKVPHGWAISLGMVEQSQAIADALEON
GWTVAIVKQSWCCFQARREGG"
complement(2558..4222)
/gene="serA"
complement(2558..4222)
/gene="serA"
/note="ORF_ID:sll1908"
/codon_start=1
/transl_table=11
/product="phosphoglycerate dehydrogenase"
/protein_id="BAAL1878.1"
/db_xref="GI:1652961"
/translation="MNLAWLQGLSLGSLSPAPALLIFRFTMAKVLVSDSIDOVGDI
ILKQAVDVKTKLSEAEITDIPEYDAIMLSATKVTETKIIOAGSOLKITIRAGVGV
INDIVPAATROGIVGVNSPEGNTIAAEHALAMMALARHIPDANKSVKESKWKQFOL
IGTEYIPKATLGVGVLGKISHVAGVAKAMKMLLAYDPIISOBRADQIGCTLVLDLIL
FSEADFIHLIPKTPETANLINAEFLAKMKPTARIINCSRGGIIDEALVTAIETAQII
GSAALDVTAQEPDLRSLREFSNVILTPHLGASTEAQVNVADVAEQIRDVLLGLPA
RVAANIGLTPDVMKRLRYLKAETLGLTVLGOLAGGRIDRLTVLCLOGDLAEYTNOSP
LVYAIAKGLSQAIRRYNVNAAIEAKERGIRVETKDSVRDYSGLSLHAKATGTMG
EHSATGALLSNEIGIRITVDDEPFINVPPNNMYMLFTLHRMPGITIGIKISGLSGSFNVI
ASMQVGRKIVRGDAIMALSDDPLPDGLLSEITKIVAGIRDAVTVKL"
4483..5205
/gene="slr1990"
4483..5205
/gene="slr1990"
/note="ORF_ID:slr1990
unknown protein"
/codon_start=1
/transl_table=11
/protein_id="BAAL1879.1"
/db_xref="GI:1652962"
/translation="MIPLLTALAQGLNFLRLRHRSGLSINPATGTMGLOAAGFLV
SVVPFIAPLVRQFVPISALATLPKVAIVMLKRPQEAAMGDDLIGFSKSWLAGAIV
WGPFWEPLVHPIEAIGLPAWGLKRGWGVIYGHLYFGLSMGTAIVDTLYFLTGILM
DHRQVYVDFEFARFVQAALAKIHTPWGAIAWMLLLALLLIGLUGGWAQLQRSPSGWA
FGGAVLSITVVDALFELWLGANNA"
5227..6240
/gene="cyaA"
5227..6240
/gene="cyaA"
/note="ORF_ID:slr1991"
/codon_start=1
/transl_table=11
/product="adenylate cyclase"
/protein_id="BAAL1780.1"
/db_xref="GI:1652963"
/translation="MDKPAPIYLVVHFSYGVDTFTPLQGRSVWTVGRSODNDLVIRDN
CISRNHAIQATEEDTEFLIDLGRSRTFVNGRRYSVPTAIDQDKITFGKTEAQFYS
DQNGTPSGLLRANPAEDWTQTNVLRHRRITVLVADMRNPTGMAQVVEELSGALQYS
PROAGHILREAGSWTKYIGDAVMAIWFHYVNEATPAETIOIILHAVNRLOAMTAKLNO
KYELFPLRIGTGNVYIGDAVMAIWFHYVNEATPAETIOIILHAVNRLOAMTAKLNO
SEKTSFYSIQDDLPQWSATVQOHTIELKQNDHPYITITGLTFPATIGTLTDHSTSLSDTTTPAS
EGP"
6281..6745
/gene="slr1992"
6281..6745

QY	428	ACTGCCACCCCAACTGATTTTCATTTAAACCCCAACCCCAACGCTT	468
Db	132420	AGTGAAGTTTGAGAATTTATTTTATGGAGCGCATGTTGCAT	132460
RESULT 6			
TETHEMOGT			
LOCUS	Tetrahymena thermophila mRNA for hemoglobin.		
DEFINITION	D13919		
ACCESSION	D13919.1 GI:217410		
VERSION	hemoglobin.		
KEYWORDS	Tetrahymena thermophila cDNA to mRNA.		
SOURCE	Tetrahymena thermophila		
ORGANISM	Eukaryota; Alveolata; Ciliophora; Oligohymenopoda; Hymenostomatida; Tetrahymenina; Tetrahymena.		
REFERENCE	1 (bases 1 to 494)		
AUTHORS	Takagi,T., Iwaasa,H., Yuasa,H., Shikama,K., Tanabe,Y.		
TITLE	Primary structure of Tetrahymena hemoglobins		
JOURNAL	Biochim. Biophys. Acta 1173 (1), 75-78 (1993)		
MEDLINE	93250050		
REFERENCE	2 (bases 1 to 494)		
AUTHORS	Takagi,T.		
TITLE	Direct Submission		
JOURNAL	Submitted (09-DEC-1992) Takashi Takagi, Biology Faculty of Science, Tohoku University; Aobayama 980, Japan (E-mail:TAKAGI@PATO.HOKU.BI.NET, Tel:022-222-1800(ex.3469), Fax:022-263-9206)		
FEATURES	Location/Qualifiers		
source	1..494		
	/organism="Tetrahymena thermophila"		
	/db_xref="taxon:5911"		
CDS	17..382		
	/codon_start=1		
	/product="hemoglobin"		
	/protein_id="BAA03014.1"		
	/db_xref="GI:418107"		
	/translation="MRKQPTVFKEGLGQAAMHAAPVLHQAQOQEDFLTMLGGPNHYKFNMAEAHKGMLNLSIIGBAKVIEHTRKDLCLK"		
BASE COUNT	175 a	74 c	85 g 160 t
ORIGIN			
Query Match	9.0%; Score 53.8; DB 3; Leng		
Best Local Similarity	49.9%; Pred. No. 0.81;		
Matches	173; Conservative 0; Mismatches 162; Ind		
QY	68	TGTTTATGACCTGGCGCGCAGAGCGATGAGCTGGCGGTGACAC	
Db	36	TTTTTGAAAGCTTGGCGGACAGCTGCTATGCATGCAGCTGTTTCCCC	
QY	128	AGTGCTGCTGACCGGAGCGTCTGCTCTTCGAGTCCCTCGACAC	
Db	96	AAGTCTTACGACATGATAGNGTCAACATTAATTAAACACATAAT	
QY	188	AGATGAAGCAGGTCAAGTTTCATGAGCTTCGTGTTGGCGGAGCAGAC	
Db	156	CTAAGTAATAGGAAGATTTTCTTACTATGCTTTTAGGAGGACCCAAATC	
QY	248	GAAGCATGTACGACGACACGCCCATCTGGTCAAGGGCCACGGCCTGG	
Db	216	AAACATGCTGAAGCTCAT-----AAAGTATGAACCTTTT	
QY	308	TTGACAAGATCAAGCAGTACCTTTGAGAGAGCGCTGCAAGAGATGGCGG	
Db	264	TTGACGCTATCATTTGAGATCTTGGCGCCACACTTAAGAATTAGGAG	
QY	368	TGATCCAGCAGCGCGGAGGTGGAGTCCACCGCGAGCAATTT	
Db	324	TTATTTGGTGAAGCAGCAAGGTTCATTGAACACACTCGTTAAAGACTGT	

CDS

1444..2178
/gene="MK0996"
/codon_start=1
/transl_table=11
/product="Archaea-specific enzyme of the ATP-grasp
superfamily"
/protein_id="AA02209.1"
/db_xref="GI:19887519"
/translation="MEVEFIRLADEIPECEFLIEGYPGIGMVGVIARHVVTECGAEP
AFNMKIDMPTAAIIVHEGRVPSAGVFLDLDSEFFYAEQADMGRPVGYNISSEVAELS
REIGVEICLAGIRAPDLEDESLYYAATSDKAAELFEGIAEPLKGTITGVSGPLL
LEGLSLGDADVLIVETPGTYPDPKAASRVVEALNEALGLSDVSKLEEAEARIAKAV
EETVORLEOQRMEEBAEPTERFV"
2240..2629
/gene="MK0997"
2240..2629
/gene="MK0997"
/codon_start=1
/transl_table=11
/product="Uncharacterized conserved protein"
/protein_id="AA02210.1"
/db_xref="GI:19887520"
/translation="MKMLAITGVSFRVVEQLLTGGSTOIIIVRSNGNLDVVRLOPQGD
VFVTVGKRELKKGVSQVATVRSVSVVHEHVVTGGDEHFEREFVAARLILNRVSGR
VREYKDCGVRREVDEMESAUSEAY"
2815..3414
/gene="MK0998"
2815..3414
/gene="MK0998"
/codon_start=1
/transl_table=11
/product="Predicted cysteine protease of the
transglutaminase-like superfamily"
/protein_id="AA02211.1"
/db_xref="GI:19887521"
/translation="MVAYQGGVLVAFKLVVYVPLTAVPPVLQIKGGEDELVVRTLA
LFVSDTHQYDVGAPPVPPVPLRAFAELELRGRWRGACGTASYMFSYLRLYGIPT
RLVLIETDGTCTVSGTGREREREYLAQVQVYVHGWMVDPVSVGLVGTSDRPVQGH
RVRPDGTLYTPGVKSYTCTQGPLVFWDLTGLVELIV"
complement(3411..5090)
/gene="cdc9"
/note="MK0999"
complement(3411..5090)
/gene="cdc9"
/codon_start=1
/transl_table=11
/product="ATP-dependent DNA ligase"
/protein_id="AA02212.1"
/db_xref="GI:19887522"
/translation="MYSSLAEEFERLERISSRKAKISLIAQLRQCPEDVVDVVALF
LANOVFGMDPRDGLGSKMRKVATATGSDSEVTELFARLGDGLTAELLKRRK
TSLDLSRLPMGEVRETFEKAIEVEGEGAVKRMRLMGLLAKRAKPEARYLVROAL
SELRTGVRESTVEEAIQAAGVSRKLVVERAHLMSNDLGLVAKVAMTKGEEGLIDLR
PMRIKPLMAQAARNVAKEALAEVGGKAVEIKLOGARVQVHSDGEVRVYTRIEDVT
HALPDIVAEKDCVDADDEFILLEGVAVINPETGKPRPFOELLHRIKRYDIEEVRKEI
PVELHLDCVLGDESIVDTPFRERRRLEIIVRECEVMVLEQVITDDPKAEAMF
HRALEMHEGVNADLDANYPGVRGKMLKVPVLETDLCVWGGINGKRRKGLG
SYLLAVDENKENLEVGKGTGMDDELRTLMFEDLIVEESGREVRFPKPVVEF
EFEDIQSPKYSYSGFALRPLRVRLDGLPEDADATIEKVRRIIEEVLQKH"
5366..6247
/gene="MK1000"
5366..6247
/gene="MK1000"
/codon_start=1
/transl_table=11
/product="Predicted serine/threonine protein kinase"
/protein_id="AA02213.1"
/db_xref="GI:19887523"
/translation="MVWVRVVQALPEMEPDFKVLRALELMRHRHWPDPRLLERT
GLDEKELGYSRLSDRMDWVTRRTQTVGLGYLQRPEDYDALRALVDQVLEGLG
PEIGVGAEADYLGISPKGAQLAVKFNRIGRITSYTKIKRYREYVKRHHISWLYNRL
TAERFEALLHLYPEGVSVPRAQNRHLVNRFEGRLEAETRVENPEAVLNVRLEE
YEHAEVGVVHGDLSQFINIVVEGDVLLIDHAWHVEVSHPSARELVERDVRNVCDFR

gene
CDS

RKYGVHRHPREFPENSQ"
6424..7323
/gene="MK1001"
6424..7323
/gene="MK1001"
/codon_start=1
/transl_table=11
/product="Ferredoxin"
/protein_id="AA02214.1"
/db_xref="GI:19887524"
/translation="MPIEIDMSCLLCEACVAACPTGAIRREDGDMNHICVGCACVKA
RGFCVCLCKMETCEPIDAIGMKGVVEPKSPPEHPDEDDVYVHPKCVGCTYCLQVCP
TDALHWPTDAEFPKAEENSKSPVYVDPDTMELKQGRKVAKIDPEKCTGCTLCAQVCP
WGATAARDVPVQSGREKNEIDKCYGCGVCAEVCPEGDLIEVDGVAKYAPEKCPACKL
CERACPVDAISINVSERGE"
7364..8158
/gene="mhpd"
/note="MK1002"
7364..8158
/gene="mhpd"
/codon_start=1
/transl_table=11
/product="2-keto-4-pentenolate hydratase hydratase"
/protein_id="AA02215.1"
/db_xref="GI:19887525"
/translation="MRVGTLYVDTGTPRPAVFLDDEVHVYDLPLVEFIEVHVSDGLGDL
DYSVPLSELRIQPPVPPPKIICFGLNYREHVEELREMGMDVSEPPMTMKAPTAVI
GHLDTVKLPREARRVDHELELAVVIGRCRKVSPEEARDAVLGTIINDVTARDIEKR
EGOWRAKSYDTFAPLGPWIELELEPDLGLEMELRVNGEVRORATDDMVRDPYELVSF
TSRVNTLEPGDIATGTPPGVGMPEPDKITEIERIGRLVHYVQ"
complement(8524..9825)
/gene="MK1003"
complement(8524..9825)
/gene="MK1003"
/codon_start=1
/transl_table=11
/product="Predicted DNA-binding protein containing a
zn-ribbon"
/protein_id="AA02216.1"
/db_xref="GI:19887526"
/translation="MDWALDDTSDPAGGCTTHAALLRAELAEAGAEVPGRPLLVRL
NPVVPKTRGNAVALPVEAPWSVDIEAVLRLKLVVRKGYPTRGCLVVCCEPPR
VCESVYEAVRRILLNPGRVKESVDDDDNVNLEGRGIVGAAVLAQRKDHVTFEG
IYAEKYGTERRVEESSIREDFRTFTFDNLDSDGDLITPMTCPVLYGVRS
VEPDLEVAQDMIKTRFVVEYEIFESNQATDAHLVRVRLADAEDYSNPVLDLTVVE
EPRIPGGHYVVRCEDEEVRVDIAAFRPARPLTEVVAALHPGDEIRVAGALRPETPK
HPTVNVYKLRVLRLEVRVNPVCGRCRSMKAGRKGFKCSGERAPEDSKIGV
EVPBELVEGTYEAPPVARRHLSKPEYLVELGLLEPSPLSR"
9852..10046
/gene="MK1004"
9852..10046
/gene="MK1004"
/codon_start=1
/transl_table=11
/product="Uncharacterized protein"
/protein_id="AA02217.1"
/db_xref="GI:19887527"
/translation="MVVGQKILASAVAHLLPFLVGVAGVYIFILTASMKDDEHLRLRGF
LIIVIGGPTSVLVSIAIVG"
10051..11028
/gene="MK1005"
10051..11028
/gene="MK1005"
/codon_start=1
/transl_table=11
/product="Predicted transcriptional regulator containing a
CHTH DNA-binding domain"
/protein_id="AA02218.1"
/db_xref="GI:19887528"
/translation="MRAELCADLEALRAGGHEVYVRIERACFDIFVTRDGRAYIVKY
LJNADGLRRREVAEBELRRTSHFLEAVPVVVALKRRHTGPLEKGVVYHVEVPVLDPLTFA
RLVGEPPKAVADRGQGVVIRADEVDELDSVRRRQLRREGGRIYLARAEADVEG

3'UTR	KFAQSTNLKSHIITHAKANNQ"
polyA_signal	1348..2330 /gene="UCRP" 2301..2306 /gene="UCRP"
BASE COUNT	635 a 552 c 601 g 542 t
ORIGIN	
Query Match	8.1%; Score 48.2; DB 10; Length 2330;
Best Local Similarity	47.8%; Pred. No. 9.1;
Matches 140;	Conservative 0; Mismatches 153; Indels 0; Gaps 0;
Oy 125	ATAAGGTGCTGGTGACCCGGAGTGTCTGCCCTTCCTCGAGTCCCTGGACATGCAAGAGC 184
Db 173	ATGATGCAGGTGCAGACCATTCCCGTGGAGACCATCGAGACACCGTGTGTGGCGGAGG 232
Oy 185	AGAAGATCAACAGCTCAAGTTTCATGACCTTCGTGTTTTGGCGGAGCACCAATACAAGG 244
Db 233	AGGAGGAGGAGCAGCAGCAGCAGGAGCGCGGGGGGGGGGACACCGCGCGGGGGGG 292
245	GCCGAAGCATGTACGAGCCACACCCCATTCTGTCAGGGCCAGCGCTGACACACCGCC 304
Ob 293	CGGCCACGGCAGCGCGGCCACCATCACACCACACCCACACCCACACCCACCGC 352
Oy 305	ACTTTGACAAGATCAAGCAGTAGTCTTGAGAGACGCTCAAGAGATGGGCTCAAGCAGG 364
Db 353	CCATGATCGCGCTGCAGCCGCTGCTGACGGAGCAGCCGACCCAAAGTCACCAACCAAG 412
Oy 365	ATCTGATCCACGCGCGCGAGTGTGGAGTCCACCCCGGACGAATTTGAC 417
Db 413	AGGTGATCTGCTGCAGACGCGGAGGAGTGGTGGCGGGGAGCAGTCGGAC 465
RESULT 11	
MUSTRANS01	
LOCUS	MUSTRANS01 . 3041 bp DNA linear ROD 27-JUL-1993
DEFINITION	Mouse delta/Yyl/NF-E1/UCRB transcription factor, exon 1.
ACCESSION	L13969
VERSION	L13969.1 GI:293843
KEYWORDS	transcription factor; zinc-finger protein.
SEGMENT	1 of 5
SOURCE	Mus musculus (library: L29SV Stratagene) 4-6 weeks liver DNA. Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE	1 (bases 1 to 3041) Safranly,G. and Perry,R.P. Characterization of the mouse gene that encodes the delta/Yyl/NF-E1/UCRB transcription factor Proc. Natl. Acad. Sci. U.S.A. 90 (12), 5559-5563 (1993)
AUTHORS	
TITLE	
JOURNAL	
MEDLINE	93296177
PUBMED	8516301
FEATURES	Location/Qualifiers 1..3041 /organism="Mus musculus" /db_xref="taxon:10090" /tissue_type="liver" /dev_stage="4-6 weeks" /tissue_lib="L29SV Stratagene" 1076..1808 1728..1737 /bound_moiety="sp1" 1776..2885 /number=1
promoter	
protein_bind	
exon	
BASE COUNT	676 a 886 c 980 g 499 t
ORIGIN	
Query Match	8.1%; Score 48.2; DB 10; Length 3041;
Best Local Similarity	47.8%; Pred. No. 8.9;
Matches 140;	Conservative 0; Mismatches 153; Indels 0; Gaps 0;
Oy 125	ATAAGGTGCTGGTGACCCGGAGTGTCTGCCCTTCCTCGAGTCCCTGGACATGCAAGAGC 184

XX XX

RESULT 3

ABL23790
ID ABL23790 standard; DNA: 9338 BP.
XX
AC ABL23790;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 22843.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ds.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
XX 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions.
XX
PS Claim 1: SEQ ID NO 22843; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Query Match 7.4%; Score 44; DB 23; Length 9338;
XX Best Local Similarity 46.0%; Pred. No. 0.22;
XX Matches 149; Conservative 0; Mismatches 175; Indels 0; Gaps 0;
XX
XX 95 GCATGAAGCTGGCGGTTCACACCTCTACGATAGGTGCTGCTGACCCGGAGCTGTGC 154
Db 7371 GCGTCTCGCTGGCTCGAAATCATGCGCAGGATGGTCAGATGTTCCCGGTGTGGCCA 7430
XX
XX 155 CTTCTTTCAGTCCCTCGACATGCAAGAGCAGAGATGAAGCTCAAGTTTCATGAGCT 214
Db 7431 ACTACTTGTAGTGTGCTCTCGCGAGGAGGAGTGTGAGTGCACGACTACGTTCACT 7490
XX
XX 215 TCGTGTGTTGGCGGAGCAGACCAATACAAAGGCGCGAGAGTGTACGACGACACGCCCATC 274
XX
Db 7491 TGGACGAGGAGTGGCGGCTTAATTCGGTGGACCGCGGTGTGGAAGAGGAGCAGCATGG 7550
XX
XX 275 TGGTCAAGGGCGACGGCTGGACACCCGACCTTTGACAAGATCAAGCAGTACCTGGAG 334
XX
Db 7551 TCGGCATCGCATGCCATCGCGGATCGACGGGTGTGATGATCTCTGGACCCAGGTACC 7610
XX
XX 335 AGACCGTCGAAGAGATGGCGCTCAAGCAGGATGTGATCCAGACCCCGCGAGTGGTGG 394

Db 7611 ATGTCAGCCGAGCGGTGACTGTATGACGATCAGAGCTATCCGCACACTGCTGAGTGT 7670
XX
Oy 395 AGTCACCCCGCAGCAATTTGACT 418
XX
Db 7671 AATCTCCGCTTAGCCTATTGTAAT 7694
XX
RESULT 4
AAZ50420
ID AAZ50420 standard; DNA: 3128 BP.
XX
AC AAZ50420;
XX
DT 18-MAY-2000 (first entry)
XX
DE Human wild type EPM2A gene encoding Laforin.
XX
KW EPM2A; laforin; chromosome 6q24; tyrosine phosphatase domain;
KW Lafora's disease; anticonvulsant; mutation; gene therapy; treatment;
KW prevention; human; ds.
XX
OS Homo sapiens.
XX
FH Key
FT CDS
FT
FT Location/Qualifiers
FT 1..996
FT /*tag= a
FT /product= "Laforin"
FT /note= "Has active catalytic site of a protein tyrosine
FT phosphatase"
FT replace (94, G)
FT /*tag= b
FT mutation
FT replace (146, G)
FT /*tag= c
FT mutation
FT replace (163, T)
FT /*tag= d
FT mutation
FT replace (179, A)
FT /*tag= e
FT mutation
FT replace (234-236, GG)
FT /*tag= f
FT /note= "G at position 235 deleted"
FT replace (322, T)
FT /*tag= g
FT mutation
FT replace (412, T)
FT /*tag= h
FT mutation
FT replace (721, T)
FT /*tag= i
FT mutation
FT replace (799..801, AAAC)
FT /*tag= j
FT /note= "A is inserted at position 800"
FT replace (836, A)
FT /*tag= k
FT mutation
FT replace (878, T)
FT /*tag= l
FT polyA_signal 3026..3031
FT /*tag= m
FT polyA_signal 3095..3100
FT /*tag= n
XX
XX WO200005405-A2.
XX
XX 03-FEB-2000.
XX
XX 20-JUL-1999; 99WO-CA00646.
XX
XX 20-JUL-1998; 98US-0093495.
XX 21-APR-1999; 99US-0130269.
XX
XX (SCHE/) SCHERER S W.
XX (MINA/) MINASSIAN B A.
XX (DELG/) DELGADO-ESCUETA A.
XX (ROUL/) ROULEAU G.

PI Scherer SW, Minasian BA, Delgado-Escueta A, Rouleau G;
 XX WPI: 2000-182718/16.
 DR P-PSDB; AAY44887.
 XX
 PT New polynucleotide encoding tyrosine phosphatase and polypeptide useful
 PT for detecting or treating Lafora's diseases and as a probe or primer to
 PT detect mutations in EPM2A gene
 XX
 XX Claim 3; Fig 13; 59pp; English.
 XX
 CC The present sequence is the human EPM2A gene encoding laforin protein.
 CC It is located on chromosome 6q24 and comprises 4 exons. Laforin
 CC comprises a tyrosine phosphatase domain associated with Lafora's
 CC disease, which is also characterised by deletions or mutations
 CC in EPM2A gene. It has anticonvulsant activity. EPM2A can be used
 CC for detecting the presence of mutations in laforin or EPM2A in samples
 CC and in gene therapy for treating or preventing Lafora's disease.
 XX
 XX Sequence 3128 BP; 876 A; 643 C; 763 G; 846 T; 0 other;
 ~

Query Match 7.1%; Score 42.6; DB 21; Length 3128;
 Best Local Similarity 50.7%; Pred. No. 0.37;
 Matches 102; Conservative 0; Mismatches 99; Indels 0; Gaps 0;

QY 1 GCGCTGCTGCACACCGCGCCACACACACAGACGCGCGGAGATCGGGGCGC 60
 DB 22 GTGGTCCACCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 81

QY 61 AGAAGCTGTTTGTATGACCTGGCGCGCGAGAGATGAAGTGGCGTTGACACCTTC 120
 DB 82 GAGCTGGGCGGTTGGGAGCGCGCGGTGCGCTCGGCTGAGGCGCGCGCGCGCGG 141

QY 121 TACCAATAGGTGCTGCTGACCGCGAGCTGCTGCCCTTCTTCAGTCCCTGGACATGCAA 180
 DB 142 GCGAGCGGGCGCTGCGCTGAGGAGCGCGCGCGCTGCTGCGTGGGAGGTGGAGCTGGCG 201

QY 181 GAGCAGAAGATGAAGCAGTGC 201
 DB 202 GCCGAGGAGCGCGCGCAGGAC 222

RESULT 5
 AAH98163
 ID AAH98163 standard; cDNA; 3134 BP.
 XX
 AC AAH98163;
 XX
 12-OCT-2001 (first entry)
 Human EST-derived coding sequence SEQ ID NO: 20.
 Human; sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse;
 KW tomato; monkey; dog; sea urchin; expressed sequence tag; EST;
 KW diagnostics; forensic test; gene mapping; genetic disorder;
 KW biodiversity; gene therapy; nutrition; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200154477-A2.
 XX
 PD 02-AUG-2001.
 XX
 PF 25-JAN-2001; 2001WO-US02687.
 XX
 PR 25-JAN-2000; 2000US-0491404.
 PR 17-JUL-2000; 2000US-0617746.
 PR 03-AUG-2000; 2000US-0631451.
 PR 15-SEP-2000; 2000US-0663870.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 XX Tang YT, Liu C, Zhou P, Qian XB, Wang Z, Chen R, Asundi V;
 PI

PI Cao Y, Drmanac RA, Zhang J, Werhman T;
 XX WPI: 2001-476164/51.
 DR P-PSDB; AAM23504.
 XX
 PT Isolated polypeptide for treatment of diseases, diagnostics, raising
 PT antibodies and research use -
 XX
 XX Claim 1; Page 195-196; 1275pp; English.
 XX
 CC The present invention provides the protein and coding sequences of novel
 CC proteins from a variety of organisms, including human, dog, cat, horse,
 CC cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea
 CC urchin and tomato. These were derived from expressed sequence tags (ESTs)
 CC from the organism of interest. They can be used in diagnostics,
 CC forensics, gene mapping, identification of mutations, to assess
 CC biodiversity and for nutritional purposes. The present sequence is a cDNA
 CC of the invention.
 XX
 XX Sequence 3134 BP; 874 A; 649 C; 759 G; 852 T; 0 other;
 ~

Query Match 7.1%; Score 42.6; DB 22; Length 3134;
 Best Local Similarity 50.7%; Pred. No. 0.37;
 Matches 102; Conservative 0; Mismatches 99; Indels 0; Gaps 0;

QY 1 GCGCTGCTGCACACCGCGCCACACACACAGACGCGCGGAGATCGGGGCGC 60
 DB 22 GTGGTCCACCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 81

QY 61 AGAAGCTGTTTGTATGACCTGGCGCGCGAGAGATGAAGTGGCGTTGACACCTTC 120
 DB 82 GAGCTGGGCGGTTGGGAGCGCGCGGTGCGCTCGGCTGAGGCGCGCGCGCGG 141

QY 121 TACCAATAGGTGCTGCTGACCGCGAGCTGCTGCCCTTCTTCAGTCCCTGGACATGCAA 180
 DB 142 GCGAGCGGGCGCTGCGCTGAGGAGCGCGCGCGCTGCTGCGTGGGAGGTGGAGCTGGCG 201

QY 181 GAGCAGAAGATGAAGCAGTGC 201
 DB 202 GCCGAGGAGCGCGCGCAGGAC 222

RESULT 6
 ABL68348
 ID ABL68348 standard; DNA; 25000 BP.
 XX
 AC ABL68348;
 XX
 DT 15-MAY-2002 (first entry)
 XX
 DE Kidney cancer related gene sequence SEQ ID NO: 6685.
 XX
 KW Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;
 KW stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;
 KW cytostatic; gene therapy; antineoplastic; Wilm's tumour; adenocarcinoma;
 KW gene; ds.
 XX
 OS Homo sapiens.
 XX
 PN WO200194629-A2.
 XX
 PD 13-DEC-2001.
 XX
 PF 30-MAY-2001; 2001WO-US10838.
 XX
 PR 05-JUN-2000; 2000US-209473P.
 PR 05-JUN-2000; 2000US-209531P.
 PR 18-SEP-2000; 2000US-233133P.
 PR 18-SEP-2000; 2000US-233617P.
 PR 20-SEP-2000; 2000US-234009P.
 PR 20-SEP-2000; 2000US-234034P.
 PR 20-SEP-2000; 2000US-234052P.
 PR 22-SEP-2000; 2000US-234509P.

PR 22-SEP-2000; 2000US-234567P.
 PR 25-SEP-2000; 2000US-234923P.
 PR 25-SEP-2000; 2000US-234924P.
 PR 25-SEP-2000; 2000US-235077P.
 PR 25-SEP-2000; 2000US-235082P.
 PR 25-SEP-2000; 2000US-235134P.
 PR 25-SEP-2000; 2000US-235280P.
 PR 26-SEP-2000; 2000US-235637P.
 PR 26-SEP-2000; 2000US-235638P.
 PR 27-SEP-2000; 2000US-235711P.
 PR 27-SEP-2000; 2000US-235720P.
 PR 27-SEP-2000; 2000US-235840P.
 PR 27-SEP-2000; 2000US-235863P.
 PR 28-SEP-2000; 2000US-236028P.
 PR 28-SEP-2000; 2000US-236032P.
 PR 28-SEP-2000; 2000US-236033P.
 PR 28-SEP-2000; 2000US-236034P.
 PR 28-SEP-2000; 2000US-236109P.
 PR 28-SEP-2000; 2000US-236111P.
 PR 29-SEP-2000; 2000US-236842P.
 PR 29-SEP-2000; 2000US-236891P.
 PR 02-OCT-2000; 2000US-237172P.
 PR 02-OCT-2000; 2000US-237173P.
 PR 02-OCT-2000; 2000US-237278P.
 PR 02-OCT-2000; 2000US-237294P.
 PR 02-OCT-2000; 2000US-237295P.
 PR 02-OCT-2000; 2000US-237316P.
 PR 03-OCT-2000; 2000US-237425P.
 PR 03-OCT-2000; 2000US-237598P.
 PR 03-OCT-2000; 2000US-237604P.
 PR 03-OCT-2000; 2000US-237606P.
 PR 03-OCT-2000; 2000US-237608P.
 PR 01-NOV-2000; 2000US-244867P.
 PR 01-NOV-2000; 2000US-245084P.
 PA (AVAL-) AVALON PHARM.
 XX Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;
 PI Soppet DR, Weaver Z;
 XX WPI; 2002-188264/24.
 XX
 PT Screening for anti-neoplastic agent involves exposing cells to a
 PT chemical agent to be tested for anti-neoplastic activity, and
 PT determining a change in expression of a gene of a signature gene set
 XX
 PS Claim 1; SEQ ID 6685; 44pp; English.
 XX
 C The present invention describes a method (M1) for screening for an
 C anti-neoplastic agent. The method involves exposing cells to a chemical
 C agent to be tested for anti-neoplastic activity, determining a change in
 C expression of at least one gene (I) of a signature gene set, where (I)
 C comprises a sequence (S) selected from 8447 sequences (given in ABL61664
 C to ABL70110), or is at least 95% identical to (S), where a change in
 C expression is indicative of anti-neoplastic activity. (I) has cytostatic
 C activity and can be used in gene therapy. M1 can be used for screening
 C an anti-neoplastic agent, and can be used for producing a product which
 C is the data collected with respect to the anti-neoplastic agent as a
 C result of M1, and the data is sufficient to convey the chemical
 C structure and/or properties of the agent. M1 can be used in the
 C treatment of cancer such as colon, breast, stomach, lung, thyroid,
 C esophageal, ovarian, kidney, prostate or pancreatic cancer,
 C adenocarcinoma, carcinoma, clear cell cancer, infiltrating ductal cancer,
 C infiltrating lobular cancer, squamous cell carcinoma, neuroendocrine
 C carcinoma, papillary carcinoma and Wilm's tumour.
 XX
 SQ Sequence 25000 BP; 5912 A; 6790 C; 6491 G; 5807 T; 0 other;
 Query Match 7.1%; Score 42.4; DB 24; Length 25000;
 Best Local Similarity 47.7%; Pred. No. 0.79;
 Matches 124; Conservative 0; Mismatches 136; Indels 0; Gaps 0;
 QY 137 CTGACCCGGAGCTGCTGCCCTCTTCGAGTCCTCCAGACATGCAAGAGAGATGAAGC 196

Db 22166 CAGACCATGTGCCACCTCTCTCCTGCACACAGGCGCCCATGATGCAGAGGAGTGAAGA 22225
 QY 197 AGGTCAAGTTTCATGAGCTTCGTGTTTGGCGAGCAGACCAATACAAAGGCCCGCAAGCATGT 256
 Db 22226 AGGACGAGACACACCGCCACCTGGAGCGCATGAAGAGAACATGGAACAGACCATTA 22285
 QY 257 AGGACGACACGCCCATCTGCTCAAGGGCCAGCGCTTGACCAACCGCCACCTTTGACAGA 316
 Db 22286 AGGACCTGCAGCACCGCGTGGACGAAGCGGAGCATGCCCTCAAGGGCGCGCAAGAAC 22345
 QY 317 TCAAGCAGTACTTGGAGAGAGCGCTCAAGAGATGGCGCTCAAGCAGCATGTGATCCAGC 376
 Db 22346 AGCTGCAGAAGCTGGAAGCGGGTGGGAGCTGGAGAAATGAGCTGGAGCCGAGCAGA 22405
 QY 377 AGCGCGCGGAGTGGTGAG 396
 Db 22406 AGCGCAACGACAGTGGGTG 22425
 RESULT 7
 AAX90823
 ID AAX90823 standard; DNA; 705 BP.
 XX
 AC AAX90823;
 XX
 DT 13-JAN-2000 (first entry)
 XX
 DE DNA encoding human Smad6 protein.
 XX
 KW Smad6 protein; expression; anticoagulant; fibrinolytic function;
 KW vascular endothelium; modulate; transforming growth factor beta; PAI-1;
 KW TGF beta; over-expression; plasminogen activator inhibitor 1; stroke;
 KW thrombomodulin; TM; atherosclerotic lesion; myocardial infarction;
 KW congestive heart failure; dilated cardiomyopathy; deep venous thrombosis;
 KW disseminated intravascular thrombosis; sepsis; trauma; septic shock;
 KW systemic inflammatory response syndrome; multiple organ dysfunction;
 KW atherosclerotic plaque rupture; anti-metastatic agent; ds.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT mat_peptide 1..705
 FT /*tag= a
 FT /product= "Smad6 protein"
 XX
 PN WO9950296-A1.
 XX
 PD 07-OCT-1999.
 XX
 PF 25-MAR-1999; 99WO-US06595.
 XX
 PR 27-MAR-1998; 98US-0079681.
 XX
 PA (ELIL) LILLY & CO ELI.
 XX
 PI Grinnell BW, Myers LJ, Richardson MA;
 XX WPI; 1999-610838/52.
 DR P-PSDB; AAY26343.
 DR
 XX Use of Smad6 and Smad7 for developing products for treating, e.g.
 PT myocardial infarction and multiple organ dysfunction syndrome
 XX
 PS Claim 1; Page 49; 60pp; English.
 XX
 CC The present sequence is a DNA encoding human Smad6 protein. Smad6
 CC mediates the expression of anticoagulant and fibrinolytic functions in
 CC the vascular endothelium, and also modulates transforming growth factor
 CC beta (TGF beta) secretion. Over-expression of Smad6 increases the
 CC secretion of plasminogen activator inhibitor 1 (PAI-1) and suppresses the
 CC expression of thrombomodulin (TM). Smad6 is predominantly expressed in
 CC atherosclerotic lesions. It can be used for treating or preventing TGF


```

XX AC AAX15147;
XX DT 26-APR-1999 (first entry)
XX DE
XX DE Nucleotide sequence of the human Smad6 cDNA.
XX KW Smad6; Smad7; Smad7 minimal promoter; TGF-beta activity;
XX KW cell growth modulation; cancer treatment; fibrosis; Smad6;
XX KW lung cancer; eye defect reduction; mammalian embryo; ss.
XX OS Homo sapiens.
XX PN W09853068-A1.
XX PD 26-NOV-1998.
XX PF 20-MAY-1998; 98WO-US10305.
XX PR 06-MAR-1998; 98US-0077033.
XX PR 20-MAY-1997; 97US-0047221.
XX PR 30-SEP-1997; 97US-0060465.
XX PR 25-FEB-1998; 98US-0075940.
XX PA (LUDW-) LUDWIG INST CANCER RES.
XX PI Heldin C, Nakao A, Ten Dijke P;
XX WI WPI; 1999-070149/06.
XX PT New isolated Smad7 nucleic acid molecules - which code for
XX PT polypeptides which inhibit transforming growth factor-beta
XX PT superfamily signalling, used to develop agents for treating, e.g.
XX PT cancer or fibrosis
XX PS Example 8; Page 73; 130pp; English.
XX CC The present sequence represents the nucleotide sequence of human Smad6
XX CC cDNA. The specification describes human and murine Smad7 sequences, and
XX CC a Smad7 minimal promoter fragment. The products can be used for
XX CC developing agents for detecting or modulating TGF-beta superfamily
XX CC activities. Such agents can be used for modulating cell growth,
XX CC e.g. to treat cancer or fibrosis. Agents which reduce the expression
XX CC or activity of Smad6 or Smad7 genes can be used for treating lung cancer
XX CC characterised by elevated expression of Smad6 or Smad7 genes. Agents
XX CC which reduce the expression or activity of a Smad7 gene or expression
XX CC product can be used for reducing eye defects in a developing mammalian
XX CC embryo.
XX Y
XX Y Sequence 1491 BP; 213 A; 546 C; 523 G; 209 T; 0 other;

Query Match 7.0%; Score 41.6; DB 20; Length 1491;
Best Local Similarity 52.7%; Pred. No. 0.53;
Matches 89; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

QY 266 ACGCCCATCTGCTCAAGGCGCCGCTGGACCCGCGCTTTCACAGATCAAGCAGT 325
DB 1076 ACCTACCTCAGGCGAGCGGCTTCTGCTGGCCAGCTCAAGCTGAGGAGCGCCGAGT 1135
QY 326 ACCTTGGAGAGACGCTGCAAGAGATGGCGCTCAACGAGATGTGATCCAGCAGCGCGC 385
DB 1136 CGGTGCGGGAGACGCGGACAGATCGGCTTGGGATCTGCTGACGAGGAGCCGAGG 1195
QY 386 GAGTGGTGGAGTCCACCGCGGACGAATTTGACTTNCCTCAACACTGCCG 434
DB 1196 GCGTGTGGGCTTACAAACCGGCGGAGCACCCCATCTTCTGTAACATCCCC 1244

RESULT 10
AAD36162
ID AAD36162 standard; DNA; 2887 BP.
XX
AC AAD36162;

```

```

XX DT 09-AUG-2002 (first entry)
XX DE Human Smad6 DNA.
XX KW Human; Smad6 protein; antisense; cardiovascular disease; infection;
XX KW inflammation; cancer; therapy; gene; ss.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
XX FT CDS 937..2427
XX FT /*tag= a
XX FT /product= "Human Smad6 protein"
XX PN W0200228878-A1.
XX PD 11-APR-2002.
XX PF 01-OCT-2001; 2001WO-US30645.
XX PR 04-OCT-2000; 2000US-0679298.
XX PA (ISIS-) ISIS PHARM INC.
XX PI Monia BP, Cowser LM;
XX WI WPI; 2002-394345/42.
XX WI P-PSDB; AAE22846.
XX PT Oligonucleotides, useful for the modulation of Smad6 expression in the
XX PT treatment or prophylaxis of e.g. cardiovascular disease, are targeted
XX PT to nucleic acid molecule encoding Smad6
XX PS Example 13; Page 95-98; 110pp; English.
XX CC The invention relates to an antisense oligonucleotide targeted to a
XX CC nucleic acid molecule encoding human Smad6 protein, which specifically
XX CC hybridises with the nucleic acid and inhibits its expression. Antisense
XX CC compounds of the invention are used for inhibiting the expression of
XX CC Smad6 in cells and tissues in the treatment of a disease or condition
XX CC associated with Smad6 such as cardiovascular disease, cancer, infection
XX CC and inflammation. They are also useful in the diagnostics, as research
XX CC reagents, in kits and in antisense therapy. The present sequence is
XX CC human Smad6 DNA.
XX SQ Sequence 2887 BP; 509 A; 899 C; 929 G; 549 T; 1 other;

Query Match 7.0%; Score 41.6; DB 24; Length 2887;
Best Local Similarity 52.7%; Pred. No. 0.65;
Matches 89; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

QY 266 ACGCCCATCTGCTCAAGGCGCCGCTGGACCCGCGCTTTCACAGATCAAGCAGT 325
DB 2012 ACCTACCTCAGGCGAGCGGCTTCTGCTGGCCAGCTCAACCTGAGGAGCGCCGAGT 2071
QY 326 ACCTTGGAGAGACGCTGCAAGAGATGGCGCTCAACGAGATGTGATCCAGCAGCGCGC 385
DB 2072 CGGTGCGGGAGACGCGGACAGATCGGCTTGGGATCTGCTGACGAGGAGCCGAGC 2131
QY 386 GAGTGGTGGAGTCCACCGCGGACGAATTTGACTTNCCTCAACACTGCCG 434
DB 2132 CGGTGTGGGCTTACAAACCGGCGGAGCACCCCATCTTCTGTAACATCCCC 2180

RESULT 11
AAT36036
ID AAT36036 standard; DNA; 3083 BP.
XX
AC AAT36036;
XX
DT 20-NOV-1996 (first entry)
XX

```


OY 266 ACGCCCATCTGCTCAAGGCCACCGCGCTGGACCAACCGGCACCTTTGACACAGATCAAGCAGT 325
 DB 1324 ACCTACCTCAGGGCAGCGGCTTCTGCTGGGCCAGCTCAACCTGGAGCAGCCGCGAGT 1383
 OY 326 ACCTTTGGAGAGCGCTGCAAGAGATGGCGTCAAGCAGGATGTGATCCAGCAGCAGCGCGCG 385
 DB 1384 CGGTGGCGAAGCGCAGCAGAGATCGGCTTCGGCATCTGCTCAGCAGGAGCCGCGAG 1443
 OY 386 GAGTGGTGGAGTCCACCGCGCAGCAATTTGACTTNCCTCAACAACTGCCG 434
 DB 1444 GCGTGTGGCGCTACAAACCGCGCAGCACCCCATCTCTGCTCAACTCCCC 1492

RESULT 13

AAV81861
 ID AAV81861 standard; cDNA: 3083 BP.

XX AC AAV81861;

XX 11-MAR-1999 (first entry)

DE Human rchd534 encoding cDNA.

KW Human; cardiovascular disease; atherosclerosis; ischaemia; restenosis;
 KW reperfusion; hypertension; arterial inflammation; diagnosis; rchd528; ds.
 XX Homo sapiens.

PH Key Location/Qualifiers
 FT CDS 1032..1739
 FT /*tag= a

XX US5849578-A.

PN 15-DEC-1998.

PD 15-MAR-1996; 96US-0616844.

PF 09-FEB-1996; 96US-0599654.

PR 10-FEB-1995; 95US-0386844.

PR 07-JUN-1995; 95US-0458873.

PR 15-MAR-1996; 96US-0616844.

XX (MILL-) MILLENNIUM PHARM INC.

PI Faib DA;

XX WPI: 1999-069743/06.

XX P-PSDB: AAW89301.

PT DNA encoding rchd528 polypeptide - associated with cardiovascular

PT disease

XX Example; Fig 34; 122pp; English.

XX The present sequence encodes rchd534 protein. The present invention
 CC describes rchd528 protein. A method has been developed for producing
 CC the rchd528 gene product. The present invention describes methods and
 CC compositions for the treatment and diagnosis of cardiovascular diseases,
 CC including: atherosclerosis; ischaemia; restenosis; reperfusion;
 CC hypertension; and arterial inflammation.

XX Sequence 3083 BP; 805 A; 770 C; 746 G; 759 T; 3 other;

Query Match

Best Local Similarity 7.0%; Score 41.6; DB 20; Length 3083;

Matches 89; Conservative 52.7%; Pred. No. 0.66;

Matches 89; Mismatches 80; Indels 0; Gaps 0;

OY 266 ACGCCCATCTGCTCAAGGCCACCGCGCTGGACCAACCGGCACCTTTGACACAGATCAAGCAGT 325

DB 1324 ACCTACCTCAGGGCAGCGGCTTCTGCTGGGCCAGCTCAACCTGGAGCAGCCGCGAGT 1383

OY 326 ACCTTTGGAGAGCGCTCAAGAGATGGCGTCAAGCAGGATGTGATCAGCAGCGCGCG 385
 DB 1384 CGGTGGCGAAGCGCGCAGCAGATCGGCTTCGGCATCTCTCAGCAGGAGCCGCGAG 1443
 OY 386 GAGTGGTGGAGTCCACCGCGCAGCAATTTGACTTNCCTCAACAACTGCCG 434
 DB 1444 GCGTGTGGCGCTACAAACCGCGCAGCACCCCATCTCTGCTCAACTCCCC 1492

RESULT 14

AAAB8579
 ID AAA88579 standard; cDNA: 3083 BP.

XX AAAB8579;

XX 05-FEB-2001 (first entry)

DE Human rchd534 cDNA.

KW Human; rchd534 gene; differential expression; HUVEC; shear stress;
 KW endothelial cell; cardiovascular disease; inflammation;
 KW atherosclerosis; antiinflammatory; antiatherosclerotic; diagnosis;
 KW gene therapy; chromosome 15; ss.

OS Homo sapiens.

XX Key Location/Qualifiers

FT CDS 1032..1739

FT /*tag= a

XX US6124433-A.

XX 26-SEP-2000.

XX 06-OCT-1997; 97US-0944496.

XX 09-FEB-1996; 96US-0599654.

XX 10-FEB-1995; 95US-0386844.

XX 07-JUN-1995; 95US-0485573.

XX (BGHM) BRIGHAM & WOMENS HOSPITAL.

XX (MILL-) MILLENNIUM PHARM INC.

XX Gimbrone MA, Faib DA;

XX WPI: 2000-611017/58.

XX P-PSDB: AAB19627.

XX Novel isolated rchd502 polypeptides, differentially expressed in
 PT response to endothelial cell shear stress, used for diagnosis,
 PT monitoring clinical trails, and treating cardiovascular diseases such
 PT as ischemia -

XX Example 9.2; Fig 34A-D; 123pp; English.

XX The present sequence is that of cDNA corresponding to the coding
 CC region and 5' and 3' regions of the novel human rchd534 gene on
 CC chromosome 15. This gene is differentially expressed (up-regulated)
 CC in response to endothelial cell shear stress, and is not induced by
 CC interleukin-1. The cDNA corresponds to clone pFCHD534, which was
 CC isolated from HUVECs subjected to laminar stress. rchd534 is a
 CC homologue of the Drosophila gene Mothers against decapentaplegic
 CC (MAD). It encodes a 235-amino acid protein (see AAB19627).
 CC up-regulation of the rchd534 gene in a disease state may reflect a
 CC protective role for the gene product in responding to disease.
 CC Alternatively, it may have a causative or exacerbating effect on
 CC the disease state. Modulation of rchd534 gene expression, or the
 CC activity of its gene product, will provide a protective effect.
 CC Knowledge of the gene and its protein product may provide for drugs
 CC with greater specificity for the treatment of inflammation and
 CC atherosclerosis, and be used to monitor clinical trials of drugs in
 CC human patients. rchd534 is 1 of 8 novel human genes of the
 CC invention (see AAA88576-83) characterised as being differentially

CC expressed in cardiovascular disease states, and which are of
CC diagnostic or therapeutic use.

XX Sequence 3083 BP; 805 A; 770 C; 746 G; 759 T; 3 other;

Query Match 7.0%; Score 41.6; DB 21; Length 3083;
Best Local Similarity 52.7%; Pred. No. 0.66;
Matches 89; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

QY 266 ACCCCCATCTGGTCAAGGGCCAGCGCTGGACCAACCGCCACTTTGACAAGATCAAGCAGT 325
DB 1324 ACCTACTCAGGCGACGGCTTCTGCTGGCGCAGCTCAACCTGGAGCAGCGCAGT 1383

QY 326 ACCTTGAGAGACGCTGCAAGAGATGGCGTCAAGCAGGATGTGATCCAGCAGCGCCG 385
DB 1384 CGGTGGGCGGACGCGCAGCAAGATCGGCTTCGGCATCTGCTCAGCAAGGAGCGCGACG 1443

QY 386 GAGTGTGGAGTCCACCGCGCAGCAATTTGACTTNCACCAACTCGGC 434
DB 1444 GCGTGGGCGCTACAACCGCGGAGCAGCCCATCTTCTGCTCAACTCCCC 1492

RESULT 15

AAZ89823 standard; cDNA: 3083 BP.

AAZ89823;

05-MAY-2000 (first entry)

Human cardiovascular disease associated gene rchd534.

Differentially expressed; cardiovascular disease; atherosclerosis;
Ischaemia; reperfusion; hypertension; restenosis; arterial inflammation;
rchd534; transmembrane protein; ss.

Homo sapiens.

US6020463-A.

01-FEB-2000.

06-OCT-1997; 97US-0944423.

09-FEB-1996; 96US-0599654.

10-FEB-1995; 95US-0386844.

07-JUN-1995; 95US-0485573.

(BGHM) BRIGHAM & WOMENS HOSPITAL.
(MILL-) MILLENNIUM PHARM INC.

Gimbrone MA, Faib DA;

WPI: 2000-146911/13.

Marker proteins for the diagnosis of cardiovascular diseases such as
atherosclerosis and hypertension, comprising peptide sequences derived
from the rchd523 transmembrane protein -

Examples; Fig 34; 121pp; English.

This sequence represents the human rchd534 gene sequence. This sequence
is related to the rchd523 transmembrane polypeptide encoded by cDNA
contained in the plasmid pfchd523. The rchd523 protein is differentially
expressed in diseased cells compared to healthy cells. The rchd523
protein may be used as a marker protein for the diagnosis of
cardiovascular diseases including atherosclerosis, ischaemia,
reperfusion, hypertension, restenosis and arterial inflammation. rchd523
peptides may be used as antigens in the production of antibodies specific
for rchd523. The anti-rchd523 antibodies may then be used in diagnostic
assays to quantitate rchd523 peptides in samples.

Sequence 3083 BP; 805 A; 770 C; 746 G; 759 T; 3 other;

Query Match 7.0%; Score 41.6; DB 21; Length 3083;
Best Local Similarity 52.7%; Pred. No. 0.66;
Matches 89; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

QY 266 ACCCCCATCTGGTCAAGGGCCAGCGCTGGACCAACCGCCACTTTGACAAGATCAAGCAGT 325
DB 1324 ACCTACTCAGGCGACGGCTTCTGCTGGCGCAGCTCAACCTGGAGCAGCGCAGT 1383

QY 326 ACCTTGAGAGACGCTGCAAGAGATGGCGTCAAGCAGGATGTGATCCAGCAGCGCCG 385
DB 1384 CGGTGGGCGGACGCGCAGCAAGATCGGCTTCGGCATCTGCTCAGCAAGGAGCGCGACG 1443

QY 386 GAGTGTGGAGTCCACCGCGCAGCAATTTGACTTNCACCAACTCGGC 434
DB 1444 GCGTGTGGGCGCTACAACCGCGGAGCAGCCCATCTTCTGCTCAACTCCCC 1492

Search completed: April 23, 2003, 12:44:53
Job time : 240 secs

Email: chauser@duke.edu.
 Location/Qualifiers
 1. 516
 /organism="Chlamydomonas reinhardtii"
 /strain="CC-1690 wild type mt+ 21gr"
 /db_xref="taxon:3055"
 /clone_lib="C. reinhardtii CC-1690, Stress II (normalized
), Lambda Zap II"
 /note="Vector: pBluescript II SK-; Site_1: EcoRI; Site_2:
 XhoI; Stress condition II library, constructed by John
 Davies and Jeffrey McDermott, combines cDNAs from CC-1690
 cells grown to mid-log phase in TAP (NH4+ - containing)
 and shifted to TAP - NO3- (24hrs); H2 production
 conditions (0, 12hr, 24hr) see Meills et al., (2000) Plant
 Phys. 122: 127-135; TAP + H2O2 (1, 12, 24 hr); TAP +
 sorbitol (1, 2, 6, 24 hr); TAP + Cd (1, 2, 6, 24 hr).
 PolyA mRNA was purified from each sample, pooled and cDNA
 synthesized. The cDNA was directionally cloned into lambda
 Zap II (Stratagene) in the EcoRI (5') and XhoI (3')
 sites. pBluescript II SK- plasmids were excised from the
 lambda Zap clones by superinfection with EXASist
 (Stratagene) phage. The library was normalized using
 method 4 described in Bonaldo et al., (1996) Genome
 Research 6: 791-806."
 93 a 181 C 147 g 95 t

zap II (Stratagene) in the EcoRI (5') and XhoI (3') sites. pBluescript II SK- plasmids were excised from the lambda ZAP clones by superinfection with ExAssist (Stratagene) phage. The library was normalized using method 4 described in Bonaldo et al., (1996) Genome Research 6: 791-806.*

BASE COUNT 93 a 181 c 147 g 95 t
ORIGIN

Query Match 19.3%; Score 115.2; DB 13; Length 516;
Best Local Similarity 59.5%; Pred. No. 5.4e-18;
Matches 195; Conservative 0; Mismatches 133; Indels 0; Gaps 0;

QY 78 CTTGGGCGGCAGAGGCCATGAAGCTGGCGGTGTACACCTTCTACCATAAGGTGCTGGC 137
| | | | | | | | | | | | | | | | | | | | | |
Db 149 CATGGGGGAGAGCCGCCCTGGAAAGGCTGTAGAGCTCTTCTATTAGGCGTATCGTTCG 208
| | | | | | | | | | | | | | | | | | | | | |
QY 138 TGACCCGGAGCTGCTGCCCTTCTTCGAGTCCCTGGACATCAAGAGCAGAGAATGAAGCA 197
| | | | | | | | | | | | | | | | | | | | | |
Db 209 GGATCCGACGCTCGCGCCCTCTTGCCCAACGTTGACATGAAGAAGCAGCGCCGCANACA 268
| | | | | | | | | | | | | | | | | | | | | |
QY 198 GGTCAAGTTTCATGACTTCGTCTTTGGCGAGCAGACCAATACAAAGGCGCGAAGCATGTA 257
| | | | | | | | | | | | | | | | | | | | | |
Db 269 GGTGCGCTTTTACTACTTACGTGTTCGGCGGCTCGGGCGCTACGAGGTCGCGACCTGGG 328
| | | | | | | | | | | | | | | | | | | | | |
QY 258 CGACGACACGCCCATCTGGTCAAGGGCCACGGCTTGGACACCGCCACCTTTCACAGAT . 317
| | | | | | | | | | | | | | | | | | | | | |
Db 329 CGCCTCCCACCGCGCGTGTATCCGGAGCAGAGGCGATGAACCAACCCACCTTTGACCTGGT 388
| | | | | | | | | | | | | | | | | | | | | |
QY 318 CAAGCAGTACCTTTGGAGAGACGCTGCAAGAGATGGGCGTCAAGCAGGATGTGTATCCAGCA 377
| | | | | | | | | | | | | | | | | | | | | |
Db 389 GGCAGCCCCACCTGGACTCCACCTGCAAGAGCTGGGGGTTGCCGAGGAGCTCAAGGCCGA 448
| | | | | | | | | | | | | | | | | | | | | |
QY 378 CGCGCGGGAGTGGTAGGTCACCGCC 405
| | | | | | | | | | | | | | | | | | | | | |
Db 449 GGC AATGGGCATCTGTGGCATCTGCCGCG 476
| | | | | | | | | | | | | | | | | | | | | |

[illegible]

JOURNAL
COMMENT
Vascular Plants. Project: 1024b
Unpublished (2001)
Contact: Charles Hauser
DCMB Box 91000
Duke University
Durham, NC 27708-1000
Tel: 919 613 8159
Fax: 919 613 8177
Email: chauser@duke.edu.

FEATURES
source
1. .521
Location/Qualifiers
/organism="Chlamydomonas reinhardtii"
/strain="CC-1690 wild type mt+ 21gr"
/db_xref="taxon:3055"
/clone_lib="C. reinhardtii CC-1690, normalized, Lambda Zap II"

/notes="Vector: pBluescript II SK-; Site.1: EcoRI; Site.2: XhoI; This library, constructed by John Davies and Jeffrey McDermott, combines cDNAs from CC-1690 cells grown to mid-log phase in TAP (acetate-containing) medium in the light, TAP medium in the dark, HS (minimal) medium in ambient levels of CO2 and HS medium bubbled with 5% CO2. PolyA mRNA was purified from each sample, pooled and cDNA synthesized. The cDNA was directionally cloned into lambda Zap II (Stratagene) in the EcoRI (5') and XhoI (3') sites. pBluescript II SK- plasmids were excised from the lambda Zap clones by superinfection with Exassist (Stratagene) phage. The library was normalized using method 4 described in Bonaldo et al (1996) Genome Research 6: 791-806."

BASE COUNT 95 a 180 c 157 g
ORIGIN

Query Match 19.3%; Score 115.2; DB 13; Length 521;
Best Local Similarity 59.5%; Pred. No. 5.4e-18;
Matches 195; Conservative 0; Mismatches 133; Indels 0; Gaps 0;

QY 78 CCTGGGGGGCGGAGAGGCATGAAGCTGGCGGTTGACACCTCTACGATAAGGTGCTGGC 137
DB 125 CATGGGGGAGAGCGCCCGTGGAAAGGCTGTAGACGCTCTTATGAGCGTATCGTTGC 184
QY 138 TGACCCGGAGCTGCTGCCCTTCTTCGAGTCCCTGGACATCAAGAGCAGGAAGATGAAGCA 197
DB 185 GGATCGCGACGCTCGCGCCCTTCTCGCCAACTGTGACATGAAGAAGCAGCGCCGCAACA 244
QY 198 GGTCAAGTTTCATGAGCTTCGTGTTGGCGGAGCAGACCAATACAAAGGCGCCAGCATGTA 257
DB 245 GTCGCGCTTTATGACTTACGTTTCGCGCGCTCGCGCGCTACGAGGCTCGGACCTGGG 304
QY 258 CGACGCACACGCCCATCTGTGTCGAAGGCCACGCGCTGGACACCGCCACTTTTGACAAAGAT 317
DB 305 CGCTCCACCGCGGCTGTATCCGAGCAGGCGCATGAACCAACACCTTTGACCTGGT 364
QY 318 CAAGCAGTACCTTTGGAGAGACGCTGCAAGAGATGGCGTCAAGCAGGATGTGATCCAGCA 377
DB 365 GGCAGGCCACCTGGACTCCACCTCTGCAAGAGCTGGGGTTGCGCAGGAGCTCAAGGCCGA 424
QY 378 CGCGCGCGGAGTGGTGGAGTCCACCGCC 405
DB 425 GGCAATGGCCATCGTGGCATCTGCCGCG 452

RESULT 4
BI527851 547 bp mRNA linear EST 29-AUG-2001
LOCUS 1024085B04.y1 C. reinhardtii CC-1690, normalized, Lambda Zap II
DEFINITION Chlamydomonas reinhardtii cDNA, mRNA sequence.

ACCESSION BI527851
VERSION BI527851.1 GI:15368425
KEYWORDS EST.
SOURCE Chlamydomonas reinhardtii.
ORGANISM Chlamydomonas reinhardtii
Eukaryota: Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
Chlamydomonadaceae; Chlamydomonas.

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

1 (bases 1 to 547)
Grossman,A., Chang,C.-W., Davies,J., Harris,E., Hauser,C., Lefebvre
P., McDermott,J.P., Shrager,J., Silflow,C. and Stern,D.
Analyses of the Chlamydomonas reinhardtii Genome: A Model,
Unicellular System for Analyzing Gene Function and Regulation in
Vascular Plants. Project: 1024b
Unpublished (2001)
Contact: Charles Hauser
DCMB Box 91000
Duke University
Durham, NC 27708-1000
Tel: 919 613 8159
Fax: 919 613 8177
Email: chauser@duke.edu.

FEATURES
source
1. .547
Location/Qualifiers

/organism="Chlamydomonas reinhardtii"
/strain="CC-1690 wild type mt+ 21gr"
/db_xref="taxon:3055"
/clone_lib="C. reinhardtii CC-1690, normalized, Lambda Zap II"
/notes="Vector: pBluescript II SK-; Site.1: EcoRI; Site.2: XhoI; This library, constructed by John Davies and Jeffrey McDermott, combines cDNAs from CC-1690 cells grown to mid-log phase in TAP (acetate-containing) medium in the light, TAP medium in the dark, HS (minimal) medium in ambient levels of CO2 and HS medium bubbled with 5% CO2. PolyA mRNA was purified from each sample, pooled and cDNA synthesized. The cDNA was directionally cloned into lambda Zap II (Stratagene) in the EcoRI (5') and XhoI (3') sites. pBluescript II SK- plasmids were excised from the lambda Zap clones by superinfection with Exassist (Stratagene) phage. The library was normalized using method 4 described in Bonaldo et al (1996) Genome Research 6: 791-806."

BASE COUNT 102 a 181 c 170 g
ORIGIN

Query Match 19.3%; Score 115.2; DB 13; Length 547;
Best Local Similarity 59.5%; Pred. No. 5.5e-18;
Matches 195; Conservative 0; Mismatches 133; Indels 0; Gaps 0;

QY 78 CCTGGGGGGCGGAGAGGCATGAAGCTGGCGGTTGACACCTCTTACGATAAGGTGCTGGC 137
DB 115 CATGGGGGAGAGCGCCCGTGGAAAGGCTGTAGACGCTCTTATGAGCGTATCGTTGC 174
QY 138 TGACCCGGAGCTGCTGCCCTTCTTCGAGTCCCTGGACATCAAGAGCAGGAAGATGAAGCA 197
DB 175 GGATCGCGACGCTCGCGCCCTTCTCGCCAACTGTGACATGAAGAAGCAGCGCCGCAACA 234
QY 198 GGTCAAGTTTCATGAGCTTCGTGTTGGCGGAGCAGACCAATACAAAGGCGCCAGCATGTA 257
DB 235 GTCGCGCTTTATGACTTACGTTTCGCGCGCTCGCGCGCTACGAGGCTCGGACCTGGG 294
QY 258 CGACGCACACGCCCATCTGTGTCGAAGGCCACGCGCTGGACACCGCCACTTTTGACAAAGAT 317
DB 295 CGCTCCACCGCGGCTGTATCCGAGCAGGCGCATGAACCAACACCTTTGACCTGGT 354
QY 318 CAAGCAGTACCTTTGGAGAGACGCTGCAAGAGATGGCGTCAAGCAGGATGTGATCCAGCA 377
DB 355 GGCAGGCCACCTGGACTCCACCTCTGCAAGAGCTGGGGTTGCGCAGGAGCTCAAGGCCGA 414
QY 378 CGCGCGCGGAGTGGTGGAGTCCACCGCC 405
DB 415 GGCAATGGCCATCGTGGCATCTGCCGCG 442

RESULT 5
BI999469 610 bp mRNA linear EST 25-OCT-2001
LOCUS 1031073H02.y2 C. reinhardtii CC-1690, Stress II (normalized),
DEFINITION Lambda Zap II Chlamydomonas reinhardtii cDNA, mRNA sequence.
ACCESSION BI999469
VERSION BI999469.1 GI:16434243

Qy	78	CTTGGGGCGCAGAAAGGCATGAAGCTGGCGTTTGACACTTCTACGATAAGGTGCTTGC	137
Db	173	CATGGGGGAGAGCGCGCTGAAAGGCTGTAGACGCTTCTATGAGCTATCTTGC	232
Qy	138	TGACCGGAGCTGCTGCCCTTCTCAGATGCCCTGGACATGCAAGACGACAGATGAAGCA	197
Db	233	GGATCCCGAGCTCGCGCCCTTCTCGCCAACTGTGACATCAAGAAGACGCGCGCAACA	292
Qy	198	GGTCAAGTTTCATGAGCTTCGTGTTTGGCGGAGCAGACCAATACAGGGCGCGAAGCATGTA	257
Db	293	GGTCGCCTTATGACTGTACGTGTTCCGGCGCTCGGGGCTCAGAGGGTCCGCACCTGGG	352
Qy	258	CGACGCACACGCCCATCTGTGTCGAAGGGCCACGGCCCTGGACCACCGCCACTTTTGACAAGAT	317
Db	353	CGCTCCACCGCGCGCTGATCCGCGAGAGGGCATGAACACACCACTTTGACCTGCT	412

Qy	258	CGAGCCACACGCCCATCTGTGTAAGGGCCACGGCCTTGACACCACCCACCTTTGACAAGAT	317
Dd	278	CGCTCCACCGCCGCCTGTATCCGCGAGCAGGGCATGAACCAACCACCTTTGACCTGGT	337
Qy	318	CAAGCAGTACTCTGGAGACGGCTGCAAGAGATGGGCGTCNAGCAGGATGTGATCCAGCA	377
Dd	338	GGCAGCCACCTGGACTCCACCCTGCAAGAGCTGGGGTTCGCGAGGAGCTCAAGGCCGA	397
Qy	378	CGCCGCGGAGTGGTGAGTCCACCCGC	405
Dd	398	GGCAATGCCCATCTGGCATCTGCCCGC	425

RESULT 8
BQ817934

LOCUS BQ817934 709 bp mRNA linear EST 01-AUG-2002

DEFINITION 1030067A09.y1 C. reinhardtii CC-1690, Deflagellation (normalized),
Lambda zap II Chlamydomonas reinhardtii cDNA, mRNA sequence.

ACCESSION BQ817934

VERSION BQ817934.1 GI:22067935

KEYWORDS EST.

SOURCE Chlamydomonas reinhardtii.
ORGANISM Chlamydomonas reinhardtii.

REFERENCE Chlamydomonas reinhardtii
AUTHORS Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
Chlamydomonadaceae; Chlamydomonas.
1 (bases 1 to 709)
Grossman,A., Chang,C.-W., Davies,J., Harris,E., Hauser,C., Lefebvre,
.P., McDermott,J.P., Shrager,J.J., Silflow,C. and Stern,D.
Analyses of the Chlamydomonas reinhardtii Genome: A Model,
Unicellular System for Analyzing Gene Function and Regulation in
Vascular Plants. Project: 1030
Unpublished (2002)

JOURNAL Contact: Charles Hauser
COMMENT DCMB Box 91000
Duke University
Durham, NC 27708-1000
Tel: 919 613 8159
Fax: 919 613 8177
Email: chauser@duke.edu.

FEATURES
source

1..709 Location/Qualifiers
/organism="Chlamydomonas reinhardtii"
/strain="CC-1690 wild type mt+ 2lgr"
/db_xref="taxon:3055"
/clone_lib="-C. reinhardtii CC-1690, Deflagellation
(normalized)", Lambda Zap II"
/notes="Vector: pBluescript II SK-; Site_1: EcoRI; Site_2:
XhoI; Deflagellation library, constructed by John Davies
and Jeffrey McDermott combines cDNAs from CC-1690 cells
which had been re-synthesizing flagella for 15, 30 and 60
min after being deflagellated by pH shock. PolyA mRNA was
purified from each sample, pooled and cDNA synthesized.
The cDNA was directionally cloned into lambda Zap II
(Stratagene) in the EcoRI (5') and XhoI (3') sites.
pBluescript II SK- plasmids were excised from the lambda
Zap clones by superinfection with EXassist (Stratagene)
phage. The library was normalized using method 4 described
in Bonaldo et al., (1996) Genome Research 6: 791-806."

BASE COUNT	129 a	219 c	231 g	130 t
ORIGIN				

Query Match 19.1%; Score 115.2; DB 14; Length 709;

Best Local Similarity 59.5%; Pred. No. 6.2e-18;

Matches 195; Conservative 0; Mismatches 133; Indels 0; Gaps 0;

Qy	78	CCTGGGGCGGCGAGAAGCATGAAGCTGGGGTTTGACACCTCTTACCATAAAGGTGCTGGC	137
Dd	93	CATGGGGGAGAGCCGCCCGTGGAAAAGCTGTAGACGTCTCTATCAGCGTATCGTGC	152
Qy	138	TGACCCGAGCTGCTGCCCTTTCTTCGAGTCCCTGGACATGCAAGAGCAGAAAGATGAAGCA	197
Dd	153	GGATCCGACGCTGCGGCCCTTCTTCGCGAACAGTGTGACATGAAGAAGAGCGCGCGCAACA	212

```

Oy 198 GGTCAAGTTCATGAGCTTCGTGTTTGGCGGAGCAGACCAATACAAAGGCCGCAAGCATGTA 257
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 213 GGTCCCTTTATGACTTACGTGTTTCGGGGCTCGGGGCTTACGAGGCTCGACCTGGG 272
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 258 CGAGCCACACGCCCATCTGTGTCAGGGCCACGGCTGGACACCGCCACTTTGACAAGAT 317
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 273 CGCTTCCACCGCCCTGTATCCGGCAGCAGGGCATGAACCAACCACTTTGACCTGGT 332
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 318 CAAGCAGTACCTTGGAGAGCGCTGCAAGAGATGGGGCTCAAGGAGGATGTGATCCAGCA 377
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 333 GGCAGCCACCTGGACTCCACCTTCAAGAGCTGGGGTTCGGCAGGAGCTCAAGGCCGA 392
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 378 CGCCCGCGGAGTGTGGAGTCCACCGC 405
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 393 GGCAATGGCCATCGTGGCATCTGCCCGC 420
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 9
961306
JUS
DEFINITION 1024076H02.y1 C. reinhardtii CC-1690, normalized, Lambda Zap II
Chlamydomonas reinhardtii cDNA, mRNA sequence.
ACCESSION BC861306
VERSION BC861306.1 GI:14242490
KEYWORDS EST.
SOURCE Chlamydomonas reinhardtii.
ORGANISM Chlamydomonas reinhardtii.
REFERENCE Grossman,A., Davies,J., Federspiel,N., Harris,E., Lefebvre,P.,
AUTHORS McDermott,J.P., Silflow,C., Stern,D. and Surzycki,R.
TITLE Analyses of the Chlamydomonas reinhardtii Genome: A Model,
Unicellular System for Analyzing Gene Function and Regulation in
Vascular Plants: project phase 2
JOURNAL Unpublished (2000)
COMMENT Contact: Charles Hauser
DCMB Box 91000
Durham, NC 27708-1000
Tel: 919 613 8159
Fax: 919 613 8177
Email: chauser@duke.edu.
FEATURES
source
1. 712
/organism="Chlamydomonas reinhardtii"
/strain="CC-1690 wild type mt+ 21gr"
/db_xref="taxon:3055"
/clone_lib="C. reinhardtii CC-1690, normalized, Lambda Zap
II".
/notes="Vector: pBluescript II SK-; Site 1: EcoRI; Site 2:
XhoI; This library, constructed by John Davies and Jeffrey
McDermott, combines cDNAs from CC-1690 cells grown to
mid-log phase in TAP (acetate-containing) medium in the
light, TAP medium in the dark, HS (minimal) medium in
ambient levels of CO2 and HS medium bubbled with 5% CO2.
polyA mRNA was purified from each sample, pooled and cDNA
synthesized. The cDNA was directionally cloned into lambda
ZAP II (Stratagene) in the EcoRI (5') and XhoI (3') sites.
pBluescript II SK- plasmids were excised from the lambda
ZAP clones by superinfection with ExAssist (Stratagene)
phage. The library was normalized using method 4 described
in Bonaldo et al (1996) Genome Research 6: 791-806."
BASE COUNT 129 a 207 c 215 g 115 t 46 others
ORIGIN
Query Match 18.6%; Score 111.2; DB 12; Length 712;
Best Local Similarity 58.6%; Pred. No. 6e-17;
Matches 191; Conservative 0; Mismatches 135; Indels 0; Gaps 0;

Oy 78 CCTGGGGCGCCAGAGCATGAGCTGGCGCTTGACACCTTCAAGTAAAGTGTCTGGC 137
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

Db 108 CATGGGGGAGAGCCGCCCTGTGAAAGGCTGTAGACGCTCTTCTATAGCGTATCGTTC 167
Oy 138 TGACCCGGAGCTGCTGCCCTTCTTCCGAGTCCCTGCACATGCAAGAGCAAGATGAAGCA 197
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 168 GGATCCGAGCTCGCGCCCTTCTTCGCAAGCTTGCATGAAGAAGCAGCGCCCAACA 227
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 198 GGTCAAGTTCATGAGCTTCGTGTTTGGCGGAGCAGACCAATACAAAGGCCGCAAGCATGTA 257
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 228 GGTCCCTTTATGACTTACGTGTTTCGGGGCTCGGGGCTTACGAGGCTCGACCTGGG 287
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 258 CGAGCCACACGCCCATCTGTGTCAGGGCCACGGCTGGACACCGCCACTTTGACAAGAT 317
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 288 CGCTTCCACCGCCCTGTATCCGGCAGCAGGGCATGAACCAACCACTTTGACCTGGN 347
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 318 CAAGCAGTACCTTGGAGAGCGCTGCAAGAGATGGGGCTCAAGCAGGATGTGATCCAGCA 377
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 348 GGCAGCCACCTGGACTCCACCTTCAAGAGCTGGGGTTCGGCAGGAGCTCAAGGCCGA 407
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Oy 378 CGCCCGCGGAGTGTGGAGTCCACCC 403
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 408 GGCAATGGCCATCGNGGCATCTGCCC 433
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 10
BI724643
LOCUS
DEFINITION 1031073H02.y1 C. reinhardtii CC-1690, Stress II (normalized),
Lambda Zap II Chlamydomonas reinhardtii cDNA, mRNA sequence.
ACCESSION BI724643
VERSION BI724643.1 GI:15700338
KEYWORDS EST.
SOURCE Chlamydomonas reinhardtii.
ORGANISM Chlamydomonas reinhardtii.
REFERENCE Grossman,A., Chang,C.-W., Davies,J., Harris,E., Hauser,C., Lefebvre
AUTHORS P., McDermott,J.P., Shrager,J., Silflow,C. and Stern,D.
TITLE Analyses of the Chlamydomonas reinhardtii Genome: A Model,
Unicellular System for Analyzing Gene Function and Regulation in
Vascular Plants, Project: 1031
JOURNAL Unpublished (2001)
COMMENT Contact: Charles Hauser
DCMB Box 91000
Durham, NC 27708-1000
Tel: 919 613 8159
Fax: 919 613 8177
Email: chauser@duke.edu.
FEATURES
source
1. 563
/organism="Chlamydomonas reinhardtii"
/strain="CC-1690 wild type mt+ 21gr"
/db_xref="taxon:3055"
/clone_lib="C. reinhardtii CC-1690, Stress II (normalized
), Lambda Zap II".
/notes="Vector: pBluescript II SK-; Site 1: EcoRI; Site 2:
XhoI; Stress condition II library, constructed by John
Davies and Jeffrey McDermott, combines cDNAs from CC-1690
cells grown to mid-log phase in TAP (NH4+ - containing)
and shifted to TAP - NO3- (24hrs); H2 production
conditions (0, 12hr, 24hr) see Melis et al., (2000) Plant
Phys. 122: 127-135; TAP + H2O2 (1, 12, 24 hr); TAP +
sorbitol (1, 2, 6, 24 hr); TAP + Cd (1, 2, 6, 24 hr).
PolyA mRNA was purified from each sample, pooled and cDNA
synthesized. The cDNA was directionally cloned into lambda
Zap II (Stratagene) in the EcoRI (5') and XhoI (3')
sites. pBluescript II SK- plasmids were excised from the
lambda Zap clones by superinfection with ExAssist
(Stratagene) phage. The library was normalized using
method 4 described in Bonaldo et al., (1996) Genome
Research 6: 791-806."
BASE COUNT 102 a 185 c 149 g 109 t 18 others

```


pBluescript II SK- plasmids were excised from the lambda ZAP clones by superinfection with ExAssist (Stratagene) phage. The library was normalized using method 4 described in Bonaldo et al., (1996) Genome Research 6: 791-806.

BASE COUNT 70 a 130 c 103 g 69 t
ORIGIN

Query Match 17.0%; Score 101.4; DB 14; Length 372;
Best Local Similarity 60.9%; Pred. No. 1.2e-14;
Matches 165; Conservative 0; Mismatches 106; Indels 0; Gaps 0;

QY 78 CTGGGGGGGAGAGGCGATGAGCTGGCGGTGACACCTTCTAGCATAGGTGCTGGC 137

DB 101 CATGGGGGAGAGCGCTCGTGAAGAGCTGTAGACGCTTCTATGAGCGGTATCGCTGC 160

QY 138 TGACCGGGAGCTGCTGCCCTTCTTCGAGTCCCTGGACATCAAGAGCAGAAGATGAAGCA 197

DB 161 GGATCCGAGCTCGCGCCCTTCTTCGCCAAGCTTGACATGAAGAGCAGCGCCGCAACA 220

QY 198 GGTCAAGTTGATGAGCTTCTGTTGGGAGGAGCAGCAACCAACAAGGCGGCAAGCATGTA 257

DB 221 GGTGGCTTTATGACTTACGTGTTCGGCGCTCGGGCGCTACGAGGGTGGCGACCTGGG 280

QY 258 CGACGACACGCCATCTGCTCAAGGCGCAGCGCTGGACCCGCCACTTTGACAGAT 317

DB 281 CGCTCCCGCGCGCTGTATCGCGAGGAGGCGATGAACACCAACCACTTTGACCTGGT 340

QY 318 CAAGCAGTACTCTGGAGAGCGCTGCAAGAG 348

DB 341 GGAGCGCCACTGGACTCCACCTGCAAGAG 371

RESULT 13

LOCUS B0822410

DEFINITION 1030100E12.y1 C. reinhardtii CC-1690, Deflagellation (normalized),

ACCESION B0822410

VERSION B0822410.1 GI:22073072

KEYWORDS EST.

SOURCE Chlamydomonas reinhardtii.

ORGANISM Chlamydomonas reinhardtii.

REFERENCE 1 (bases 1 to 350)

AUTHORS Grossman,A., Chang,C.-W., Davies,J., Harris,E., Hauser,C., Lefebvre

TITLE Analyses of the Chlamydomonas reinhardtii Genome: A Model,

JOURNAL Unpublished (2002)

COMMENT Contact: Charles Hauser

DCMB Box 91000

Duke University

Durham, NC 27708-1000

Tel: 919 613 8159

Fax: 919 613 8177

Email: chauser@duke.edu.

FEATURES

Source

1..350

/organism="Chlamydomonas reinhardtii"

/strain="CC-1690 wild type mt+ 21gr"

/db_xref="taxon:3055"

/clone_lib="C. reinhardtii CC-1690, Deflagellation

(normalized), Lambda Zap II"

/note="Vector: pBluescript II SK-; Site_1: EcoRI; Site_2:

XhoI; Deflagellation library, constructed by John Davies

pBluescript II SK- plasmids were excised from the lambda ZAP clones by superinfection with ExAssist (Stratagene) phage. The library was normalized using method 4 described in Bonaldo et al., (1996) Genome Research 6: 791-806.

BASE COUNT 65 a 120 c 100 g 65 t
ORIGIN

Query Match 15.8%; Score 94.2; DB 14; Length 350;
Best Local Similarity 62.6%; Pred. No. 6.8e-13;
Matches 147; Conservative 0; Mismatches 88; Indels 0; Gaps 0;

QY 78 CTGGGGGGGAGAGGCGATGAGCTGGCGGTGACACCTTCTAGCATAGGTGCTGGC 137

DB 112 CATGGGGGAGAGCGCGCTGGAAGAGCTGTAGACGCTTCTATGAGCGGTATCGCTGC 171

QY 138 TGACCGGGAGCTGCTGCCCTTCTTCGAGTCCCTGGACATCAAGAGCAGAAGATGAAGCA 197

DB 172 GGATCCGAGCTCGCGCCCTTCTTCGCCAAGCTTGACATGAAGAGCAGCGCCGCAACA 231

QY 198 GGTCAAGTTGATGAGCTTCTGTTGGGAGGAGCAGCAACCAACAAGGCGGCAAGCATGTA 257

DB 232 GGTGGCTTTATGACTTACGTGTTCGGCGCTCGGGCGCTACGAGGGTGGCGACCTGGG 291

QY 258 CGACGACACGCCATCTGCTCAAGGCGCAGCGCTGGACCCGCCACTTTGAC 312

DB 292 CGCTCCCGCGCGCTGTATCGCGAGGAGGCGATGAACCAACCACTTTGAC 346

RESULT 14

LOCUS AV638207

DEFINITION AV638207 Chlamydomonas reinhardtii 5% CO2 Chlamydomonas reinhardtii

ACCESION AV638207.1 GI:10781527

VERSION EST.

KEYWORDS Chlamydomonas reinhardtii.

SOURCE Chlamydomonas reinhardtii.

ORGANISM Chlamydomonas reinhardtii.

REFERENCE 1 (bases 1 to 303)

AUTHORS Asamizu,E., Miura,K., Kuchio,K., Inoue,Y., Fukuzawa,H., Ohya,K.,

Nakamura,Y., and Tabata,S.

TITLE Generation of expressed sequence tags from low-CO2 and high-CO2

JOURNAL adapted cells of Chlamydomonas reinhardtii

MEDLINE DNA Res. 7 (5), 305-307 (2000)

COMMENT 20539644

Contact: Erika Asamizu

The First Laboratory for Plant Gene Research

Kazusa DNA Research Institute

Yana 1532-3, Kisarazu, Chiba 292-0812, Japan

Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.

FEATURES

Source

1..303

/organism="Chlamydomonas reinhardtii"

/strain="C9"

/db_xref="taxon:3055"

/clone_lib="Chlamydomonas reinhardtii 5% CO2"

/note="Vector: pBluescript II SK-; Site_1: EcoRI; Site_2:

XhoI; The cDNA library was constructed from cells cultured

in a medium with bubbling air containing 5% carbon

dioxide"

BASE COUNT

ORIGIN

Query Match

Best Local Similarity

Matches

Conservative

Mismatches

Indels

Gaps

0;

QY 78 CTGGGGGGGAGAGGCGATGAGCTGGCGGTGACACCTTCTAGCATAGGTGCTGGC 137

DB 125 CATGGGGGAGAGCGCGCTGGAAGAGCTGTAGACGCTTCTATGAGCGGTATCGCTGC 184

Result No.	Query			DB	ID	Description
	Score	Match	Length			
1	42	7.0	1058	4	US-09-452-239-11	Sequence 11, Appl
2	41.6	7.0	1280	4	US-09-096-776B-4	Sequence 4, Appl
3	41.6	7.0	1491	4	US-09-082-092-9	Sequence 9, Appl
4	41.6	7.0	1524	4	US-08-840-767-3	Sequence 3, Appl
5	41.6	7.0	1817	4	US-09-288-292A-45	Sequence 45, Appl
6	41.6	7.0	3083	2	US-08-480-994-36	Sequence 36, Appl
7	41.6	7.0	3083	2	US-08-616-844-36	Sequence 36, Appl
8	41.6	7.0	3083	2	US-08-599-654-36	Sequence 36, Appl
9	41.6	7.0	3083	2	US-08-485-573-36	Sequence 36, Appl
10	41.6	7.0	3083	3	US-08-944-868A-36	Sequence 36, Appl
11	41.6	7.0	3083	3	US-08-944-423A-36	Sequence 36, Appl
12	41.6	7.0	3083	3	US-08-925-743-36	Sequence 36, Appl
13	41.6	7.0	3083	3	US-08-944-496-36	Sequence 36, Appl
14	41.6	7.0	3083	4	US-08-925-767-36	Sequence 36, Appl
15	41.6	7.0	3084	3	US-08-826-246-11	Sequence 11, Appl
16	41.6	7.0	3084	3	US-08-944-495-11	Sequence 11, Appl
17	41.6	7.0	3084	3	US-09-126-640-6	Sequence 6, Appl
18	41.6	7.0	3084	4	US-08-925-588-11	Sequence 11, Appl
19	41.6	7.0	3084	4	US-09-288-292A-6	Sequence 6, Appl
20	39.65	7.0	7218	1	US-08-232-463-14	Sequence 14, Appl
21	38.4	6.4	423	1	US-08-470-179-186	Sequence 186, Appl
22	38.2	6.4	1341	4	US-09-199-637A-133	Sequence 133, Appl
23	38.2	6.4	2176	6	5320958-1	Patent No. 5320958
24	38	6.4	3234	1	US-08-264-534-31	Sequence 31, Appl
25	38	6.4	3234	1	US-08-083-590A-10	Sequence 10, Appl
26	38	6.4	3234	1	US-08-465-500-31	Sequence 31, Appl
27	38	6.4	3234	2	US-08-346-128-31	Sequence 31, Appl

```

: RESULT 2
: US-09-096-776B-4
: Sequence 4, Application US/09096776B
: Patent No. 6270994
: GENERAL INFORMATION:
: APPLICANT: Miyazono, Kohei
: APPLICANT: Kawabata, Masahiro
: TITLE OF INVENTION: SMAD6 AND USES THEREOF
: FILE REFERENCE: L0461/7038
: CURRENT APPLICATION NUMBER: US/09/096, 776B
: CURRENT FILING DATE: 1998-06-12

```

```

; PRIOR APPLICATION NUMBER: US 60/049,990
; PRIOR FILING DATE: 1997-06-13
; PRIOR APPLICATION NUMBER: US 60/053,040
; PRIOR FILING DATE: 1997-07-18
; PRIOR APPLICATION NUMBER: US 60/066,173
; PRIOR FILING DATE: 1997-11-18
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 1280
; TYPE: DNA.
; ORGANISM: Homo sapiens
US-09-096-776B-4

```

	Query Match	7.0%;	Score 41.6;	DB 4;	Length 1280;
	Best Local Similarity	52.7%;	Pred. No. 0.066;		
	Matches 89; Conservative	0;	Mismatches 80;	Indels	0;
	Gaps				
QY	266	ACGCCCATCTGGTCAAGGGCCACGCGCTGGACCACACCGCCACTTTGACAAGATCAACGACGT	325		
	405	ACCTACTCAGGCAGCGGGCTTC TGCTGGGCCAGGCTCAAACCTGGAGCAGCGACGGAGT	464		
QY	326	ACCTTGGAGAGCGCTGCAAGAGATGGCGCTCAAGCAGGATGTGTATCCACGACGCGCGCCG	385		
Dd	465	CGTGCGCGGAACGCGCACCAAGATCGGCTTCGCGCATCTGCTCAGCAAGGACCCGACG	524		
QY	386	GAGTGGTGGAGTCCACCCGCGGACGAATTGTACTTNCCCACAACTGCGC	434		
Dd	525	GCGTGTGGGCTTACAACCGCGGGAGCACGCCCATCTTCGTCAACTCCCC	573		

RESULT 3
US-09-082-092-9
: Sequence 9, Application US/09082092
: Patent No. 6251628
: GENERAL INFORMATION:
: APPLICANT: Nakao, Atsuhito
: APPLICANT: Moren, Anita
: APPLICANT: Heuchel, Rainer
: APPLICANT: Itoh, Susumu
: APPLICANT: Afrahkhte, Mozghan
: APPLICANT: Souchehlnytskyi, Serhiy
: APPLICANT: Brodin, Greger
: APPLICANT: Landstrom, Marene
: APPLICANT: Heldin, Nils-Erik
: APPLICANT: Heldin, Carl-Henrik
: APPLICANT: ten Dijke, Peter
: TITLE OF INVENTION: SMAD7 AND USES THEREOF
: NUMBER OF SEQUENCES: 15
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Wolf, Greenfield & Sacks, P.C.
: STREET: 600 Atlantic Avenue
: CITY: Boston
: STATE: MA
: COUNTRY: U.S.A.
: Zip: 02210-2211
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: DOS
: SOFTWARE: FASTSEO for Windows Version 2.0
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/082,092
: FILING DATE: 20-MAY-1998
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 60/047,221
: FILING DATE: 20-MAY-1997
: APPLICATION NUMBER: 60/060,465
: FILING DATE: 30-SEP-1997
: APPLICATION NUMBER: 60/075,940
: FILING DATE: 25-FEB-1998
: APPLICATION NUMBER: 60/077,033

```

: FILING DATE: 06-MAR-1998
: ATTORNEY/AGENT INFORMATION:
: NAME: Van Amsterdam, John R.
: REGISTRATION NUMBER: 40,212
: REFERENCE/DOCKET NUMBER: L0461/7032
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 617-720-3500
: TELEFAX: 617-720-2441
: TELEX:
: INFORMATION FOR SEQ ID NO: 9:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1491 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: US-09-082-092-9

```

	Query Match	7.0%	Score 41.6;	DB 4;	Length 1491;	
	Best Local Similarity	52.7%;	Pred. No. 0.069;			
	Matches	89;	Conservative	0; Mismatches	80; Indels	0; Gaps
QY	266	ACGCCCATCTCGTCAAGGGCCACGGCCTGGACCACCGGCACCTTTGACACAGATCAAGCACT	325			
Db	1076	ACCTACCTTCAGGGCAGCGGCTTCGCTTGCGGCAGCTCAACTGGAGCACGCCAGCGACT	1135			
QY	326	ACCTTGGAGAGACGCTGCAAGAGATGGCGGTCAAGCAGGATGTGATCAGACACGCCGCCG	385			
Db	1136	CGGTGCGCGNAACGGCAGCAAGATCGGCTTCGGGATCCTGCTCAGCAGGAGACCGCGAG	1195			
QY	386	GAGTGGTGGAGTCCACCGCGACGAATTGTGACTTNCCCAACAACATGCGCG	434			
Db	1196	GCFTGTGGCGCTACAAACCGCGCGAGCACCACCATCTTCGTCAACTCCCG	1244			

```

RESULT 4
US-08-840-767-3
; Sequence 3, Application US/08840767B
; Patent No. 6255464
; GENERAL INFORMATION:
; APPLICANT: Vogelstein, Bert
; APPLICANT: Kinzler, Kenneth W.
; APPLICANT: Riggins, Gregory J.
; APPLICANT: Thibault, Sam
; TITLE OF INVENTION: MAD-Related Genes In the Human
; FILE REFERENCE: 01107 05548
; CURRENT APPLICATION NUMBER: US/08/840,767B
; CURRENT FILING DATE: 1997-04-16
; EARLIER APPLICATION NUMBER: 60/015,823
; EARLIER FILING DATE: 1996-04-18
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: FASTSEQ for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 1524
; TYPE: DNA
; ORGANISM: Homo sapiens
US-08-840-767-3

```

[illegible]

RESULT 5

US-09-288-292A-45

```
; Sequence 45, Application US/09288292A
; Patent No. 6359194
; GENERAL INFORMATION:
; APPLICANT: Dean A. Falb
; APPLICANT: Katherine Galvin
; APPLICANT: Michael Donovan
; APPLICANT: Dennis Huszar
; APPLICANT: Michael A. Gimbrown, Jr.
; TITLE OF INVENTION: Compositions and Methods for the Treatment and Diagnosis of
; FILE OF INVENTION: Cardiovascular Disease
; FILE REFERENCE: 7853-140-999
; CURRENT APPLICATION NUMBER: US/09/288,292A
; CURRENT FILING DATE: 1999-04-08
; PRIOR APPLICATION NUMBER: 08/870,434
; PRIOR FILING DATE: 1997-06-06
; PRIOR APPLICATION NUMBER: 08/799,910
; PRIOR FILING DATE: 1997-02-13
; PRIOR APPLICATION NUMBER: 60/011,787
; PRIOR FILING DATE: 1996-02-16
; PRIOR APPLICATION NUMBER: 08/485,573
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 08/386,844
; PRIOR FILING DATE: 1995-02-10
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 45
; LENGTH: 1817
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-288-292A-45
```

```
Query Match 7.0%; Score 41.6; DB 4; Length 1817;
Best Local Similarity 52.7%; Pred. No. 0.074;
Matches 89; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

QY 266 ACGCCCATCTGTGTCAGAGGCGCCGCTGGACCGCCACTTTGACAGATCAAGCAGT 325
DB 1230 ACCTACCTCAGGCGAGCGGCTTCTGCTGGCGCAGCTCAACCTGGAGCAGCGCAGT 1289

QY 326 ACCTTGGAGAGCGCTGCAAGAGATGGCGCTCAACAGATCTGTATCCAGCAGCGCGC 385
DB 1290 CGGTGCGGCGAAGCGCGAGAGATCGGCTTGGCATCTGCTCAGCAGGAGCGCGCAG 1349

QY 386 GAGTGGTGGAGTCCACCGCGCAGCAATTTGACTTNCACCAACTGCCG 434
; 1350 CGGTGCGGCTTACAAACCGCGCGAGCAGCCCATCTTCTCAACTCCCC 1398
```

RESULT 6

US-08-480-994-36

```
; Sequence 36, Application US/08480994
; Patent No. 5834248
; GENERAL INFORMATION:
; APPLICANT: FALB, DEAN A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; FILE OF INVENTION: TREATMENT AND DIAGNOSIS OF CARDIOVASCULAR DISEASE
; NUMBER OF SEQUENCES: 38
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
```

```
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/480,994
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/485,573
; FILING DATE: 07-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/386,844
; FILING DATE: 10-FEB-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7853-033
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864
; INFORMATION FOR SEQ ID NO: 36:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3083 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: unknown
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 16
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 30
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 2911
; US-08-480-994-36

Query Match 7.0%; Score 41.6; DB 2; Length 3083;
Best Local Similarity 52.7%; Pred. No. 0.086; 80; Indels 0; Gaps 0;
Matches 89; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

QY 266 ACGCCCATCTGTGTCAGAGGCGCCGCTGGACCGCCACTTTGACAGATCAAGCAGT 325
DB 1324 ACCTACCTCAGGCGAGCGGCTTCTGCTGGCGCAGCTCAACCTGGAGCAGCGCAGT 1383

QY 326 ACCTTGGAGAGCGCTGCAAGAGATGGGGTCAACAGAGTGTGTATCCAGCAGCGCGCG 385
DB 1384 CGGTGCGGCGAAGCGCGAGAGATCGGCTTGGCATCTGCTCAGCAGGAGCGCGCAG 1443

QY 386 GAGTGGTGGAGTCCACCGCGCAGCAATTTGACTTNCACCAACTGCCG 434
DB 1444 CGGTGCGGCTTACAAACCGCGCGAGCAGCCCATCTTCTCAACTCCCC 1492
```

RESULT 7

US-08-616-844-36

```
; Sequence 36, Application US/08616844
; Patent No. 5849578
; GENERAL INFORMATION:
; APPLICANT: FALB, DEAN A.
; TITLE OF INVENTION: COMPOSITION AND METHODS FOR THE
; FILE OF INVENTION: TREATMENT AND DIAGNOSIS OF CARDIOVASCULAR DISEASE
; NUMBER OF SEQUENCES: 54
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
```

```

RESULT 8
US-08-599-654-36
; Sequence 36, Application US/08599654
; Patent No. 5882925
; GENERAL INFORMATION:
; APPLICANT: FALB, DEAN A
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; TITLE OF INVENTION: TREATMENT AND DIAGNOSIS OF CARDIOVASCULAR DISEASE
; NUMBER OF SEQUENCES: 54
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711

```

```

RESULT 9
US-08-485-573-36
; Sequence 36, Application US/08485573
; Patent No. 5968770
; GENERAL INFORMATION:
; APPLICANT: FALB, DEAN A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; TREATMENT AND DIAGNOSIS OF CARDIOVASCULAR DISEASE
; NUMBER OF SEQUENCES: 38
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA

```

ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/485,573
FILING DATE: 07-JUN-1995
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/386,844
FILING DATE: 10-FEB-1995
ATTORNEY/AGENT INFORMATION:
NAME: Coruzzi, Laura A.
REGISTRATION NUMBER: 30,742
REFERENCE/DOCKET NUMBER: 7853-032
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 36:
SEQUENCE CHARACTERISTICS:
LENGTH: 3083 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: unknown
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: misc_feature
LOCATION: 16
FEATURE:
NAME/KEY: misc_feature
LOCATION: 30
FEATURE:
NAME/KEY: misc_feature
LOCATION: 2911
US-08-485-573-36

Query Match 7.0%; Score 41.6; DB 2; Length 3083;
Best Local Similarity 52.7%; Pred. No. 0.086;
Matches 89; Conservative 0; Mismatches 80; Indels 0; Gaps 0;
QY 266 ACGCCCATCTGTCAGGCGCCAGCGCTGGACACCGCCACTTTGACAAAGATCAAGCAGT 325
Db 1324 ACCTACCTCAGGCGAGCGGCTTCTGCTGGCCAGCTCAACCTGGAGCAGCGCAGCAGT 1383
QY 326 ACCTTGGAGAGCGCTGCAAGAGATGGCGTCAAGCAGCATGTGATCCAGCAGCGCGCG 385
b 1384 CGTGCGGCGAAGCGCCAGCAAGATCGGCTTCGGCATCTGCTCAGCAAGAGCGCGCAGC 1443
QY 386 GAGTGTGAGTCCACCGCGGACGAATTTGACTTNCACCAACTGCGC 434
Db 1444 GCGTGTGGCCTACAAACCGCGGAGCAGCCCATCTTCGTCAACTCCCC 1492

RESULT 10
US-08-944-868A-36
Sequence 36, Application US/08944868A
Patent No. 6018025
GENERAL INFORMATION:
APPLICANT: FALB, DEAN A
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
TREATMENT AND DIAGNOSIS OF CARDIOVASCULAR DISEASE
NUMBER OF SEQUENCES: 54
CORRESPONDENCE ADDRESS:
ADDRESSEE: PENNIE & EDMONDS
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036-2711
COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/944,868A
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/599,654
FILING DATE:
APPLICATION DATA:
APPLICATION NUMBER: US 08/386,844
FILING DATE: 10-FEB-1995
ATTORNEY/AGENT INFORMATION:
NAME: CORUZZI, LAURA A
REGISTRATION NUMBER: 30,742
REFERENCE/DOCKET NUMBER: 7853-041
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 36:
SEQUENCE CHARACTERISTICS:
LENGTH: 3083 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: unknown
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: misc_feature
LOCATION: 16
FEATURE:
NAME/KEY: misc_feature
LOCATION: 30
FEATURE:
NAME/KEY: misc_feature
LOCATION: 2911
US-08-944-868A-36

Query Match 7.0%; Score 41.6; DB 3; Length 3083;
Best Local Similarity 52.7%; Pred. No. 0.086;
Matches 89; Conservative 0; Mismatches 80; Indels 0; Gaps 0;
QY 266 ACGCCCATCTGTCAGGCGCCAGCGCTGGACACCGCCACTTTGACAAAGATCAAGCAGT 325
Db 1324 ACCTACCTCAGGCGAGCGGCTTCTGCTGGCCAGCTCAACCTGGAGCAGCGCAGCAGT 1383
QY 326 ACCTTGGAGAGCGCTGCAAGAGATGGCGTCAAGCAGCATGTGATCCAGCAGCGCGCG 385
Db 1384 CGTGCGGCGAAGCGCCAGCAAGATCGGCTTCGGCATCTGCTCAGCAAGAGCGCGCAGC 1443
QY 386 GAGTGTGAGTCCACCGCGGACGAATTTGACTTNCACCAACTGCGC 434
Db 1444 GCGTGTGGCCTACAAACCGCGGAGCAGCCCATCTTCGTCAACTCCCC 1492

RESULT 11
US-08-944-423A-36
Sequence 36, Application US/08944423A
Patent No. 6020463
GENERAL INFORMATION:
APPLICANT: FALB, DEAN A
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
TREATMENT AND DIAGNOSIS OF CARDIOVASCULAR DISEASE
NUMBER OF SEQUENCES: 54
CORRESPONDENCE ADDRESS:
ADDRESSEE: PENNIE & EDMONDS
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036-2711

STATE: New York

COUNTRY: USA
 ZIP: 10036-2711
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent In Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/944,496
 FILING DATE: 06-OCT-1997
 CLASSIFICATION: 514
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/599,654
 FILING DATE: 09-FEB-1996
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/485,573
 FILING DATE: 07-JUN-1995
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/386,844
 FILING DATE: 10-FEB-1995
 ATTORNEY/AGENT INFORMATION:
 NAME: CORUZZI, LAURA A
 REGISTRATION NUMBER: 30,742
 REFERENCE/DOCKET NUMBER: 7853-104
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (212) 790-9090
 TELEFAX: (212) 869-8864
 TELEX: 66141 PENNIE
 INFORMATION FOR SEQ ID NO: 36:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 3083 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: both
 TOPOLOGY: unknown
 MOLECULE TYPE: cDNA
 FEATURE:
 NAME/KEY: misc_feature
 LOCATION: 16
 FEATURE:
 NAME/KEY: misc_feature
 LOCATION: 30
 FEATURE:
 NAME/KEY: misc_feature
 LOCATION: 2911
 US-08-944-496-36

Query Match 7.0%; Score 41.6; DB 3; Length 3083;
 Best Local Similarity 52.7%; Pred. No. 0.086;
 Matches 89; Conservative 0; Mismatches 80; Indels 0; Gaps 0;
 QY 266 ACCCCCATCTGTCAGAGCGCCGCTGGACCAACCCGCTTTGACAAAGATCAAGCAGT 325
 DB 1324 ACCTACTCAGGCGAGCGGCTTCTGCTGGCGGAGCTCAACCTGGAGCAGCGCAGT 1383
 QY 326 ACCTTGAGAGCGCTCAAGAGATGGCGTCAAGAGGATGTGTCACCAACTGCGC 305
 DB 1384 CGGTGGCGGAACCGCGAGCAAGATCGGCTTCGGCATCTCGCTCAGCAGGAGCGCCGACG 1443
 QY 386 GAGTGGTGGAGTCCACCCGCGAGCAATTTGACTTNCACCAACTGCGC 434
 DB 1444 GCGTGGGGCTACACCGCGGAGCAGCCCATCTTCGTCACACTCCCC 1492

RESULT 14
 US-08-925-767-36
 Sequence 36, Application US/08925767
 Patent No. 6225084
 GENERAL INFORMATION:
 APPLICANT: FALB, DEAN A.
 TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
 TREATMENT AND DIAGNOSIS OF CARDIOVASCULAR DISEASE
 NUMBER OF SEQUENCES: 38
 CORRESPONDENCE ADDRESS:

ADDRESSEE: PENNIE & EDMONDS
 STREET: 1155 Avenue of the Americas
 CITY: New York
 STATE: New York
 COUNTRY: USA
 ZIP: 10036-2711
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent In Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/925,767
 FILING DATE: 09-SEPT-1997
 CLASSIFICATION: 514
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/485,573
 FILING DATE: 07-JUN-1995
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/386,844
 FILING DATE: 10-FEB-1995
 ATTORNEY/AGENT INFORMATION:
 NAME: CORUZZI, LAURA A.
 REGISTRATION NUMBER: 30,742
 REFERENCE/DOCKET NUMBER: 7853-097
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (212) 790-9090
 TELEFAX: (212) 869-8864
 TELEX: 66141 PENNIE
 INFORMATION FOR SEQ ID NO: 36:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 3083 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: both
 TOPOLOGY: unknown
 MOLECULE TYPE: cDNA
 FEATURE:
 NAME/KEY: misc_feature
 LOCATION: 16
 FEATURE:
 NAME/KEY: misc_feature
 LOCATION: 30
 FEATURE:
 NAME/KEY: misc_feature
 LOCATION: 2911
 US-08-925-767-36

Query Match 7.0%; Score 41.6; DB 4; Length 3083;
 Best Local Similarity 52.7%; Pred. No. 0.086;
 Matches 89; Conservative 0; Mismatches 80; Indels 0; Gaps 0;
 QY 266 ACCCCCATCTGTCAGAGCGCCGCTGGACCAACCCGCTTTGACAAAGATCAAGCAGT 325
 DB 1324 ACCTACTCAGGCGAGCGGCTTCTGCTGGCGGAGCTCAACCTGGAGCAGCGCAGT 1383
 QY 326 ACCTTGAGAGCGCTCAAGAGATGGCGTCAAGAGGATGTGTCACCAACTGCGC 385
 DB 1384 CGGTGGCGGAACCGCGAGCAAGATCGGCTTCGGCATCTCGCTCAGCAGGAGCGCCGACG 1443
 QY 386 GAGTGGTGGAGTCCACCCGCGAGCAATTTGACTTNCACCAACTGCGC 434
 DB 1444 GCGTGGGGCTACACCGCGGAGCAGCCCATCTTCGTCACACTCCCC 1492

RESULT 15
 US-08-826-246-11
 Sequence 11, Application US/08826246
 Patent No. 6048709
 GENERAL INFORMATION:
 APPLICANT: Falb, Dean
 TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
 THE TREATMENT AND DIAGNOSIS OF
 TITLE OF INVENTION: CARDIOVASCULAR DISEASE

```

: NUMBER OF SEQUENCES: 44
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: PENNIE & EDMONDS LLP
: STREET: 1155 Avenue of the Americas
: CITY: New York
: STATE: NY
: COUNTRY: USA
: ZIP: 10036-2711
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: DOS
: SOFTWARE: FastSeq Version 2.0
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/826,246
: FILING DATE: 28-MAR-1997
: CLASSIFICATION: 800
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/799,910
: FILING DATE: 13-FEB-1997
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 60/011,787
: FILING DATE: 16-FEB-1996
: ATTORNEY/AGENT INFORMATION:
: NAME: Coruzzi, Laura A
: REGISTRATION NUMBER: 30,742
: REFERENCE/DOCKET NUMBER: 7853-078-999
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (212)7909090
: TELEFAX: (212)8699741
: TELEX: 66141 PENNIE
: INFORMATION FOR SEQ ID NO: 11:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 3084 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: both
: TOPOLOGY: linear
: MOLECULE TYPE: cdna
: FEATURE:
: NAME/KEY: Coding Sequence
: LOCATION: 1032...1736
: OTHER INFORMATION:
:
: US-08-826-246-11

```

```

Query Match      7.0%; Score 41.6; DB 3; Length 3084;
Best Local Similarity 52.7%; Pred. No. 0.086;
Matches 89; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

Qy 266 ACGCCCATCTGGTCAAGGCGCAGCGCTGGACGACGCCCACTTTGACAAGATCAAGCAGT 325
      || || || || || || || || || || || || || || || || || || || ||
1324 ACCTACCTCAGGCGCGGCTTCTGCTGGCGCCAGCTCAACCTGGAGCAGCGCAGCAGT 1383

Qy 326 ACCTTGGAGAGCGCTGCAAGAGATGGCGTCAAGCAGGATGTGATCCAGCAGCGCGCGG 385
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1384 CGGTGCGGCGACGCGCAGCAGATCGGCTTCGGCATCTGCTCAGCAGGAGCGCGACG 1443

Qy 386 GAGTGTGGAGTCCACCGCGAGATTTCACCTTCCCAACAACTGCGG 434
      ||||| || || || || || || || || || || || || || || || || ||
Db 1444 GCCTGTGGCCTACAAACCGCGGAGCAGCCCATCTTCGTCAACTCCCC 1492

```

Search completed: April 23, 2003, 13:50:28
Job time : 58 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 23, 2003, 12:35:20 ; Search time 103 seconds
(without alignments)
6317.485 Million cell updates/sec

Title: US-09-920-953-2

Perfect score: 598

Sequence: 1 gcgtgcctgcagcgcggc.....agccatttgcaccaaagcc 598

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 709820 seqs, 544064369 residues

total number of hits satisfying chosen parameters: 1419640

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published_Applications_NA:

- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
- 2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq.*
- 3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq.*
- 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq.*
- 5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
- 6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq.*
- 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
- 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq.*
- 9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
- 10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq.*
- 11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq.*
- 12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq.*
- 13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq.*
- 14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	42	7.0	1058	10	US-09-452-239-11
2	41.6	7.0	1280	10	US-09-923-922-4
3	41.6	7.0	1280	10	US-09-934-456-1590
4	41.6	7.0	1817	10	US-09-924-417-64
5	41.6	7.0	3083	9	US-09-970-820-36
6	41.6	7.0	3083	9	US-09-985-718-36
7	41.6	7.0	3083	10	US-09-371-900-36
8	41.6	7.0	3083	10	US-09-924-417-62
9	41.6	7.0	3084	8	US-08-825-486-11
10	41.6	7.0	3084	8	US-08-870-434-6
11	41.6	7.0	3084	9	US-09-560-150-6
12	41.6	7.0	3084	10	US-09-372-044-11
13	41	6.9	353	10	US-09-864-761-17688
14	41	6.9	355	10	US-09-864-761-25057
15	41	6.9	582	10	US-09-864-761-8326
16	39.4	6.6	417	9	US-09-918-995-3955
17	38.2	6.4	1341	9	US-09-975-719-133
18	38.2	6.4	2176	8	US-08-808-031A-1
19	38	6.4	1566	10	US-09-815-242-7758

Sequence 1, Appli	20	38	6.4	7332	10	US-09-944-849-1	Sequence 1, Appli
Sequence 1, Appli	21	38	6.4	7673	10	US-09-815-925-1	Sequence 1, Appli
Sequence 130, App	c 22	37.8	6.3	1470	10	US-09-073-009-130	Sequence 130, App
Sequence 130, App	c 23	37.8	6.3	1470	10	US-09-793-306-130	Sequence 130, App
Sequence 15, Appli	24	37.2	6.2	3388	10	US-09-866-582-15	Sequence 15, Appli
Sequence 219, App	25	37	6.2	859	12	US-10-062-254-219	Sequence 219, App
Sequence 6, Appli	26	36.8	6.2	1467	10	US-09-981-900B-6	Sequence 6, Appli
Sequence 57, Appli	27	36.8	6.2	2412	12	US-10-002-600-57	Sequence 57, Appli
Sequence 569, App	28	36.8	6.2	2412	12	US-10-044-090-569	Sequence 569, App
Sequence 7888, Ap	29	36.8	6.2	2838	10	US-09-815-242-7888	Sequence 7888, Ap
Sequence 47, Appli	30	36.8	6.2	3773	10	US-09-925-302-47	Sequence 47, Appli
Sequence 53, Appli	31	36.4	6.1	1527	10	US-09-729-674-53	Sequence 53, Appli
Sequence 4, Appli	32	36.4	6.1	1852	10	US-09-969-852-4	Sequence 4, Appli
Sequence 4059, Ap	33	36.2	6.1	1596	10	US-09-815-242-4059	Sequence 4059, Ap
Sequence 6, Appli	34	36.2	6.1	3624	9	US-09-988-462-6	Sequence 6, Appli
Sequence 5, Appli	35	36.2	6.1	4509	9	US-10-124-800-5	Sequence 5, Appli
Sequence 27, Appli	36	36.2	6.1	4512	9	US-10-124-800-27	Sequence 27, Appli
Sequence 37947, A	37	36	6.0	493	9	US-09-918-995-37947	Sequence 37947, A
Sequence 1021, Ap	38	36	6.0	1384	10	US-09-867-550-1021	Sequence 1021, Ap
Sequence 756, App	c 39	36	6.0	3665	10	US-09-925-300-756	Sequence 756, App
Sequence 99, Appli	40	35.8	6.0	4233	9	US-09-991-496-99	Sequence 99, Appli
Sequence 99, Appli	41	35.8	6.0	4233	10	US-09-874-923-99	Sequence 99, Appli
Sequence 100, App	42	35.8	6.0	4917	9	US-09-991-496-100	Sequence 100, App
Sequence 100, App	43	35.8	6.0	4917	10	US-09-874-923-100	Sequence 100, App
Sequence 98, Appli	44	35.8	6.0	4929	9	US-09-991-496-98	Sequence 98, Appli
Sequence 98, Appli	45	35.8	6.0	4929	10	US-09-874-923-98	Sequence 98, Appli

ALIGNMENTS

RESULT 1

US-09-452-239-11
; Sequence 11, Application US/09452239
; Patent No. US20020081693A1
; GENERAL INFORMATION:
; APPLICANT: Rafalski, Antoni J.
; APPLICANT: Pader, Gary M.
; APPLICANT: Cahoon, Rebecca E.
; TITLE OF INVENTION: Plant Caffeyol-CoA O-Methyltransferase
; FILE REFERENCE: BB1284 US NA
; CURRENT APPLICATION NUMBER: US/09/452,239
; CURRENT FILING DATE: 1999-12-01
; EARLIER APPLICATION NUMBER: 60/110,594
; EARLIER FILING DATE: 1998-December-02
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 11
; LENGTH: 1058
; TYPE: DNA
; ORGANISM: Oryza sativa
US-09-452-239-11

Query Match 7.0%; Score 42; DB 10; Length 1058;

Best Local Similarity 52.9%; Pred No. 0.0021;

Matches 90; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

QY 162 CGAGTCCCTGGACATGCAAGAGCAGCAGAGTCAAGTTCATGAGCTTCGTGT 221

Db 619 CAACATACCTCACTACCACGAGCGGTGATGAAGTTCGAGGCGGCGCTCGCG 678

QY 222 TGGCGGAGCAGACCAATACAGGGCGGAAGCATGTACGACGACAGCCCATCTGTC 281

Db 679 CTACGACACACAGCTGTGAGACGGCTCGTCTGCCCGGCGGCGGCGGCGGCA 738

QY 282 GGGCCAGCGGCTTGACACCGCCACTTTGACAAGATCAAGCAGTACTCTTG 331

Db 739 GTACATCCCTACTACCGGCGACTTCTGCTCGAGCTCAACAAGGCGGCTCG 788

RESULT 2

US-09-923-922-4
; Sequence 4, Application US/09923922

Patent No. US20020106732A1
GENERAL INFORMATION:
APPLICANT: Miyazono, Kohei
APPLICANT: Kawabata, Masahiro
TITLE OF INVENTION: SHAD6 AND USES THEREOF
FILE REFERENCE: L0461/7120
CURRENT APPLICATION NUMBER: US/09/923,922
PRIOR FILING DATE: 2001-08-07
PRIOR APPLICATION NUMBER: US/09/66,776
PRIOR FILING DATE: 1998-06-12
PRIOR APPLICATION NUMBER: US 60/049,990
PRIOR FILING DATE: 1997-06-13
PRIOR APPLICATION NUMBER: US 60/053,040
PRIOR FILING DATE: 1997-07-18
PRIOR APPLICATION NUMBER: US 60/066,173
PRIOR FILING DATE: 1997-11-18
NUMBER OF SEQ ID NOS: 10
SOFTWARE: PatentIn version 3.0
SEQ ID NO 4
LENGTH: 1280
TYPE: DNA
ORGANISM: Homo sapiens
S-09-923-922-4

Query Match 7.0%; Score 41.6; DB 10; Length 1280;
Best Local Similarity 52.7%; Pred. No. 0.003;
Matches 89; Conservative 0; Mismatches 80; Indels 0; Gaps 0;
QY 266 ACCCCCATCTGGTCAAGGCGCAGCGCTGGACACCGCCACTTTTGACAAGATCAAGCAGT 325
DB 405 ACCTACCTCAGGCGAGCGGCTTCTGCTGGCGCAGCTCAACCTGGAGCGCGCAGT 464
QY 326 ACCTTGGAGAGCGCTGCAAGAGATGGCGCTCAAGCAGGATGTGATCCCAACAACGCGC 385
DB 465 CGGTGGCGGCGACCGCCACCAAGATCGGCTTCGGCATCTCTCAGCAAGAGCGCGAGC 524
QY 386 GAGTGGTGGAGTCCACCGCGCAGCAATTTGACTTNCCTCAACAACGCGC 434
DB 525 CGGTGGCGGCTTACAACCGCGCGAGCAGCCCATCTTCTGTCAACTCCCC 573

RESULT 3
US-09-954-456-1590
Sequence 1590, Application US/09954456
Patent No. US20020115057A1
GENERAL INFORMATION:
APPLICANT: Young, Paul
TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using Can
FILE REFERENCE: 689290-76
CURRENT APPLICATION NUMBER: US/09/954,456
PRIOR FILING DATE: 2001-09-18
PRIOR APPLICATION NUMBER: US/60/233,617
PRIOR FILING DATE: 2000-09-18
PRIOR APPLICATION NUMBER: US/60/234,052
PRIOR FILING DATE: 2000-09-20
PRIOR APPLICATION NUMBER: US/60/234,923
PRIOR FILING DATE: 2000-09-25
PRIOR APPLICATION NUMBER: US/60/235,134
PRIOR FILING DATE: 2000-09-25
PRIOR APPLICATION NUMBER: US/60/235,637
PRIOR FILING DATE: 2000-09-26
PRIOR APPLICATION NUMBER: US/60/235,638
PRIOR FILING DATE: 2000-09-26
PRIOR APPLICATION NUMBER: US/60/235,711
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: US/60/235,720
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: US/60/235,840
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: US/60/235,863
PRIOR FILING DATE: 2000-09-27
NUMBER OF SEQ ID NOS: 2276

SOFTWARE: PatentIn version 3.0
SEQ ID NO 1590
LENGTH: 1280
TYPE: DNA
ORGANISM: Homo sapiens
US-09-954-456-1590
Query Match 7.0%; Score 41.6; DB 10; Length 1280;
Best Local Similarity 52.7%; Pred. No. 0.003;
Matches 89; Conservative 0; Mismatches 80; Indels 0; Gaps 0;
QY 266 ACCCCCATCTGGTCAAGGCGCAGCGCTGGACACCGCCACTTTTGACAAGATCAAGCAGT 325
DB 405 ACCTACCTCAGGCGAGCGGCTTCTGCTGGCGCAGCTCAACCTGGAGCGCGCAGT 464
QY 326 ACCTTGGAGAGCGCTGCAAGAGATGGCGCTCAAGCAGGATGTGATCCCAACAACGCGC 385
DB 465 CGGTGGCGGCGACCGCCACCAAGATCGGCTTCGGCATCTCTCAGCAAGAGCGCGAGC 524
QY 386 GAGTGGTGGAGTCCACCGCGCAGCAATTTGACTTNCCTCAACAACGCGC 434
DB 525 CGGTGGCGGCTTACAACCGCGCGAGCAGCCCATCTTCTGTCAACTCCCC 573

RESULT 4
US-09-924-417-64
Sequence 64, Application US/09924417
Patent No. US20020142441A1
GENERAL INFORMATION:
APPLICANT: Falb, Dean
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE TREATMENT AND DIAGNOSIS OF CARDIOVASCULAR DISEASE
NUMBER OF SEQUENCES: 67
CORRESPONDENCE ADDRESS:
ADDRESSEE: PENNIE & EDMONDS LLP
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/924,417
FILING DATE: 07-Aug-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/034,286
FILING DATE: 04-MAR-1998
APPLICATION NUMBER: 08/870,434
FILING DATE: 06-JUN-1997
APPLICATION NUMBER: 08/799,910
FILING DATE: 13-FEB-1997
APPLICATION NUMBER: 60/011,787
FILING DATE: 16-FEB-1996
APPLICATION NUMBER: 08/599,654
FILING DATE: 09-FEB-1996
APPLICATION NUMBER: 08/485,573
FILING DATE: 07-JUN-1995
APPLICATION NUMBER: 08/386,844
FILING DATE: 10-FEB-1995
ATTORNEY/AGENT INFORMATION:
NAME: Coruzzi, Laura A
REGISTRATION NUMBER: 30,742
REFERENCE/DOCKET NUMBER: 7853-114-999
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212)7909090
TELEFAX: (212)8699741
TELEX: 66141 PENNIE

INFORMATION FOR SEQ ID NO: 64:
SEQUENCE CHARACTERISTICS:
LENGTH: 1817 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: unknown
MOLECULE TYPE: cdna
SEQUENCE DESCRIPTION: SEQ ID NO: 64:
US-09-924-417-64

Query Match
Best Local Similarity 7.0%; Score 41.6; DB 10; Length 1817;
Matches 89; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

QY 266 ACGCCATCTGTCAGGCGCCAGCGCTCGACACCGCCACCTTTGACAGATCAAGCAGT 325
DB 1230 ACTTACTCAGGCGCGGCTTCTGCTGGCGCAGCTCAACCTGGAGCAGCGCAGGAGT 1289
QY 326 ACCTTGGAGAGCGCTGCAAGAGATGGCGCTCAAGCAGGATGTGATCCAGCAGCGCGCG 385
DB 1290 CGGTGGCGGCAAGCGCCACAGATCGGCTTCGGCATCTCTGCTCAGCAAGGAGCGCGAGC 1349
QY 386 GAGTGTGAGTCCACCGCGCAGCAATTTGACTTNCCTCAACAACTGCGC 434
DB 1350 GCGTGTGGGCTTACAAACCGCGGAGCAGCCCATCTTCGTCAACTCCCC 1398

RESULT 5
US-09-970-820-36
Sequence 36, Application US/09970820
Patent No. US20020170077A1
GENERAL INFORMATION:
APPLICANT: FALB, DEAN A.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
TREATMENT AND DIAGNOSIS OF CARDIOVASCULAR DISEASE
NUMBER OF SEQUENCES: 38
CORRESPONDENCE ADDRESS:
ADDRESSEE: PENNIE & EDMONDS
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/970,820
FILING DATE: 05-OCT-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/386,844
FILING DATE: 10-FEB-1995
ATTORNEY/AGENT INFORMATION:
NAME: Coruzzi, Laura A.
REGISTRATION NUMBER: 30,742
REFERENCE/DOCKET NUMBER: 7853-032
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 36:
SEQUENCE CHARACTERISTICS:
LENGTH: 3083 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: unknown
MOLECULE TYPE: cdna
FEATURE:
NAME/KEY: misc_feature
LOCATION: 16

FEATURE:
NAME/KEY: misc_feature
LOCATION: 30
FEATURE:
NAME/KEY: misc_feature
LOCATION: 2911
SEQUENCE DESCRIPTION: SEQ ID NO: 36:
US-09-970-820-36

Query Match
Best Local Similarity 7.0%; Score 41.6; DB 9; Length 3083;
Matches 89; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

QY 266 ACGCCATCTGTCAGGCGCCAGCGCTCGACACCGCCACCTTTGACAGATCAAGCAGT 325
DB 1324 ACTTACTCAGGCGCGGCTTCTGCTGGCGCAGCTCAACCTGGAGCAGCGCAGGAGT 1383
QY 326 ACCTTGGAGAGCGCTGCAAGAGATGGCGCTCAAGCAGGATGTGATCCAGCAGCGCGCG 385
DB 1384 CGGTGGCGGCAAGCGCCACAGATCGGCTTCGGCATCTCTGCTCAGCAAGGAGCGCGAGC 1443
QY 386 GAGTGTGAGTCCACCGCGCAGCAATTTGACTTNCCTCAACAACTGCGC 434
DB 1444 GCGTGTGGGCTTACAAACCGCGGAGCAGCCCATCTTCGTCAACTCCCC 1492

RESULT 6
US-09-986-718-36
Sequence 36, Application US/09986718
Patent No. US20020178458A1
GENERAL INFORMATION:
APPLICANT: FALB, DEAN A.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
TREATMENT AND DIAGNOSIS OF CARDIOVASCULAR DISEASE
NUMBER OF SEQUENCES: 38
CORRESPONDENCE ADDRESS:
ADDRESSEE: PENNIE & EDMONDS
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/986,718
FILING DATE: 09-NOV-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/485,573
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Coruzzi, Laura A.
REGISTRATION NUMBER: 30,742
REFERENCE/DOCKET NUMBER: 7853-032
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 36:
SEQUENCE CHARACTERISTICS:
LENGTH: 3083 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: unknown
MOLECULE TYPE: cdna
FEATURE:
NAME/KEY: misc_feature
LOCATION: 16

```

; NAME/KEY: misc_feature
; LOCATION: 30
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 2911
; SEQUENCE DESCRIPTION: SEQ ID NO: 36:
US-09-986-718-36

Query Match      7.0%; Score 41.6; DB 9; Length 3083;
Best Local Similarity 52.7%; Pred. No. 0.0041;
Matches 89; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

OY 266 ACGCCCTCTGGTCAAGGCCACGCCCTGGACCCGACCTTTGACAGATCAAGCAGT 325
      || || || || || || || || || || || || || || || || || || ||
DB 1324 ACCTACCTCAGGGCAGCGGCTTCTGCCCTGGGCCAGCTCAACCTGGAGCAGCGCAGT 1383
      || || || || || || || || || || || || || || || || || || ||
OY 326 ACCTTGGAGAGACGCTGCAAGAGATGGCGCTCAAGCAGGATGTGATCCAGCAGCGCGCG 385
      || || || || || || || || || || || || || || || || || || ||
DB 1384 CGGTGCGGCGAAGCGGCGAGCAGATCGGCTTGGGCATCTCTGCTAGCAAGAGCGCCGACG 1443
      || || || || || || || || || || || || || || || || || || ||
      386 GAGTGGTGGAGTCCACCGCGACGAAATTTGACTTNCCTCAACAACTGCCG 434
      || || || || || || || || || || || || || || || || || || ||
DB 1444 CGGTGTGGGCTTACAACCGCGGCGAGCACCCCATCTTCTGTCAACTCCCC 1492
      || || || || || || || || || || || || || || || || || || ||

RESULT 7
US-09-371-900-36
; Sequence 36, Application US/09371900
; Patent No. US20020137700A1
; GENERAL INFORMATION:
; APPLICANT: FALB, DEAN A
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; TREATMENT AND DIAGNOSIS OF CARDIOVASCULAR DISEASE
; NUMBER OF SEQUENCES: 54
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/371,900
; FILING DATE: 11-Aug-1999
; CLASSIFICATION: <unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/599,654
; FILING DATE: 09-FEB-1996
; APPLICATION NUMBER: US 08/485,573
; FILING DATE: 07-JUN-1995
; APPLICATION NUMBER: US 08/386,844
; FILING DATE: 10-FEB-1995
; NAME: CORUZZI, LAURA A
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7853-104
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 36:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3083 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: unknown
; MOLECULE TYPE: CDNA
; FEATURE:
```

```

; NAME/KEY: misc_feature
; LOCATION: 16
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 30
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 2911
; SEQUENCE DESCRIPTION: SEQ ID NO: 36:
US-09-371-900-36

Query Match      7.0%; Score 41.6; DB 10; Length 3083;
Best Local Similarity 52.7%; Pred. No. 0.0041;
Matches 89; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

OY 266 ACGCCCTCTGGTCAAGGCCACGCCCTGGACCCGACCTTTGACAGATCAAGCAGT 325
      || || || || || || || || || || || || || || || || || || ||
DB 1324 ACCTACCTCAGGGCAGCGGCTTCTGCCCTGGGCCAGCTCAACCTGGAGCAGCGCAGT 1383
      || || || || || || || || || || || || || || || || || || ||
OY 326 ACCTTGGAGAGACGCTGCAAGAGATGGCGCTCAAGCAGGATGTGATCCAGCAGCGCGCG 385
      || || || || || || || || || || || || || || || || || || ||
DB 1384 CGGTGCGGCGAAGCGGCGAGCAGATCGGCTTGGGCATCTCTGCTAGCAAGAGCGCCGACG 1443
      || || || || || || || || || || || || || || || || || || ||
      386 GAGTGGTGGAGTCCACCGCGACGAAATTTGACTTNCCTCAACAACTGCCG 434
      || || || || || || || || || || || || || || || || || || ||
DB 1444 CGGTGTGGGCTTACAACCGCGGCGAGCACCCCATCTTCTGTCAACTCCCC 1492
      || || || || || || || || || || || || || || || || || || ||

RESULT 8
US-09-924-417-62
; Sequence 62, Application US/09924417
; Patent No. US20020142441A1
; GENERAL INFORMATION:
; APPLICANT: Falb, Dean
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
; THE TREATMENT AND DIAGNOSIS OF CARDIOVASCULAR
; DISEASE
; NUMBER OF SEQUENCES: 67
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/924,417
; FILING DATE: 07-Aug-2001
; CLASSIFICATION: <unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/034,286
; FILING DATE: 04-MAR-1998
; APPLICATION NUMBER: 08/870,434
; FILING DATE: 06-JUN-1997
; APPLICATION NUMBER: 08/799,910
; FILING DATE: 13-FEB-1997
; APPLICATION NUMBER: 60/011,787
; FILING DATE: 16-FEB-1996
; APPLICATION NUMBER: 08/599,654
; FILING DATE: 09-FEB-1996
; APPLICATION NUMBER: 08/485,573
; FILING DATE: 07-JUN-1995
; APPLICATION NUMBER: 08/386,844
; FILING DATE: 10-FEB-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Cotuzzi, Laura A
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7853-114-999
```



```

Matches 89; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

QY 266 AGCCCATCTGTCAAGGCGCCACGGCTGGACACCGCCACTTTGACAAGATCAAGCAGT 325
  || || || || || || || || || || || || || || || || || || || || ||
Db 1324 ACCTACCTCAGGCGAGCGGCTTCTGCTGGCCAGCTCAACCTGGAGCAGCGCAGGAT 1383

QY 326 ACCTTGGAGAGACGCTGCAAGAGATGGCGTCAAGCAGGATGTGATCCAGCAGCGCGCG 385
  || || || || || || || || || || || || || || || || || || || || ||
Db 1384 CGGTGGCGGCAACGCGCAGCAAGATCGGCTTCGGCATCTGCTCAGCAAGAGCGCGCAG 1443

QY 386 GAGTGGTGGAGTCCACCCGCGACGAATTTGACTTNCCTCAACAACTGCGC 434
  || || || || || || || || || || || || || || || || || || || || ||
Db 1444 GCGTGGGCGCTACAAACCGCGGAGCAGCCCATCTTGTCAACTCCCC 1492

RESULT 11
US-09-560-150-6
; Sequence 6, Application US/09560150
; Publication No. US20030073076A1
; GENERAL INFORMATION:
; APPLICANT: FALB, Dean A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; FILE OF INVENTION: TREATMENT AND DIAGNOSIS OF CARDIOVASCULAR DISEASE
; FILE REFERENCE: 7853-126
; CURRENT APPLICATION NUMBER: US/09/560,150
; CURRENT FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 09/126,640
; PRIOR FILING DATE: 1998-07-30
; PRIOR APPLICATION NUMBER: 08/870,434
; PRIOR FILING DATE: 1997-06-06
; PRIOR APPLICATION NUMBER: 08/799,910
; PRIOR FILING DATE: 1997-02-13
; PRIOR APPLICATION NUMBER: 60/011,787
; PRIOR FILING DATE: 1996-02-16
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 3084
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-560-150-6

Query Match
Best Local Similarity 7.0%; Score 41.6; DB 9; Length 3084;
Matches 89; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

QY 266 AGCCCATCTGTCAAGGCGCCACGGCTGGACACCGCCACTTTGACAAGATCAAGCAGT 325
  || || || || || || || || || || || || || || || || || || || || ||
Db 1324 ACCTACCTCAGGCGAGCGGCTTCTGCTGGCCAGCTCAACCTGGAGCAGCGCAGGAT 1383

QY 326 ACCTTGGAGAGACGCTGCAAGAGATGGCGTCAAGCAGGATGTGATCCAGCAGCGCGCG 385
  || || || || || || || || || || || || || || || || || || || || ||
Db 1384 CGGTGGCGGCAACGCGCAGCAAGATCGGCTTCGGCATCTGCTCAGCAAGAGCGCGCAG 1443

QY 386 GAGTGGTGGAGTCCACCCGCGACGAATTTGACTTNCCTCAACAACTGCGC 434
  || || || || || || || || || || || || || || || || || || || || ||
Db 1444 GCGTGGGCGCTACAAACCGCGGAGCAGCCCATCTTGTCAACTCCCC 1492

RESULT 12
US-09-372-044-11
; Sequence 11, Application US/09372044A
; Patent No. US20020102603A1
; GENERAL INFORMATION:
; APPLICANT: Dean FALB et al.
; TITLE OF INVENTION: Compositions and Methods for the
; FILE OF INVENTION: Treatment and Diagnosis of Cardiovascular Disease
; FILE REFERENCE: 7853-152
; CURRENT APPLICATION NUMBER: US/09/372,044A
; CURRENT FILING DATE: 1999-08-11
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11

```

```

; LENGTH: 3084
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1032)...(1736)
US-09-372-044-11

Query Match
Best Local Similarity 7.0%; Score 41.6; DB 10; Length 3084;
Matches 89; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

QY 266 AGCCCATCTGTCAAGGCGCCACGGCTGGACACCGCCACTTTGACAAGATCAAGCAGT 325
  || || || || || || || || || || || || || || || || || || || || ||
Db 1324 ACCTACCTCAGGCGAGCGGCTTCTGCTGGCCAGCTCAACCTGGAGCAGCGCAGGAT 1383

QY 326 ACCTTGGAGAGACGCTGCAAGAGATGGCGTCAAGCAGGATGTGATCCAGCAGCGCGCG 385
  || || || || || || || || || || || || || || || || || || || || ||
Db 1384 CGGTGGCGGCAACGCGCAGCAAGATCGGCTTCGGCATCTGCTCAGCAAGAGCGCGCAG 1443

QY 386 GAGTGGTGGAGTCCACCCGCGACGAATTTGACTTNCCTCAACAACTGCGC 434
  || || || || || || || || || || || || || || || || || || || || ||
Db 1444 GCGTGGGCGCTACAAACCGCGGAGCAGCCCATCTTGTCAACTCCCC 1492

RESULT 13
US-09-864-761-17688/C
; Sequence 17688, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL
; FILE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aesomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30

```


; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 17688
; LENGTH: 353
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AL049829.2
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.7e+02
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.4
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.63
; OTHER INFORMATION: SWISSPROT HIT: P13533, EVALUE 2.00e-38
; OTHER INFORMATION: NT HIT: 220656.1, EVALUE 0.00e+00
; OTHER INFORMATION: EST_HUMAN HIT: C05089.1, EVALUE 0.00e+00
US-09-864-761-17688

Query Match
Best Local Similarity 6.9%; Score 41; DB 10; Length 353;
Matches 110; Conservative 0; Mismatches 115; Indels 0; Gaps 0;

QY 172 GACATCAAGAGCAGCAAGATGAAGCAGGTCAAGTTTCATGAGCTTCGTTTGGCGGAGCA 231
DB 172 GACATCAAGAGCAGCAAGATGAAGCAGGTCAAGTTTCATGAGCTTCGTTTGGCGGAGCA 231
QY 273 GCCATGATGCAGAGCAGGTGAAGAGGAGCAGCAGCCACCTGGAGCCGATG 214
DB 273 GCCATGATGCAGAGCAGGTGAAGAGGAGCAGCAGCCACCTGGAGCCGATG 214
QY 232 GACCAATACAAGGCCGAGCATGTACGAGCAGCAGCCCATCTGTCTAAGGCCACGGC 291
DB 232 GACCAATACAAGGCCGAGCATGTACGAGCAGCAGCCCATCTGTCTAAGGCCACGGC 291
QY 213 AAGAAGAATGAGCAGCAGCATTAAAGGACCTGCAGCAGCGGTGGACGAGCCGAGCAG 154
DB 213 AAGAAGAATGAGCAGCAGCATTAAAGGACCTGCAGCAGCGGTGGACGAGCCGAGCAG 154
QY 292 CTGGACCAAGCAGGTGATGATCCAGCAGCAGCGCGGAGTGTGGAG 396
DB 292 CTGGACCAAGCAGGTGATGATCCAGCAGCAGCGCGGAGTGTGGAG 396
QY 153 ATGCGCCCTCAAGGGAGGCAAGAGCAGCTGCAGAAAGCTGGAAGCGCGGTGGAGGAGTG 94
DB 153 ATGCGCCCTCAAGGGAGGCAAGAGCAGCTGCAGAAAGCTGGAAGCGCGGTGGAGGAGTG 94
QY 352 GCGGTGAGCTGAGCGCCGAGCAGAGCAGCGCGGAGTGTGGAG 49
DB 352 GCGGTGAGCTGAGCGCCGAGCAGAGCAGCGCGGAGTGTGGAG 49

RESULT 14
US-09-864-761-25057/c
; Sequence 25057, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aomic-X-1
; CURRENT APPLICATION NUMBER: US 09/864,761
; PRIOR FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 25057
; LENGTH: 355
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AL049829.3
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.5e+02
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 7.6
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 5.2
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 4.9
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 7.3
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 6.9
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 5.6
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 7
; OTHER INFORMATION: NT HIT: 220656.1, EVALUE 0.00e+00
; OTHER INFORMATION: EST_HUMAN HIT: C05089.1, EVALUE 0.00e+00
; OTHER INFORMATION: SWISSPROT HIT: P13533, EVALUE 2.00e-38
US-09-864-761-25057

Query Match
Best Local Similarity 6.9%; Score 41; DB 10; Length 355;
Matches 110; Conservative 0; Mismatches 115; Indels 0; Gaps 0;

QY 172 GACATCAAGAGCAGCAAGATGAAGCAGGTCAAGTTTCATGAGCTTCGTTTGGCGGAGCA 231
DB 172 GACATCAAGAGCAGCAAGATGAAGCAGGTCAAGTTTCATGAGCTTCGTTTGGCGGAGCA 231
QY 273 GCCATGATGCAGAGCAGGTGAAGAGGAGCAGCAGCCACCTGGAGCCGATG 214
DB 273 GCCATGATGCAGAGCAGGTGAAGAGGAGCAGCAGCCACCTGGAGCCGATG 214
QY 232 GACCAATACAAGGCCGAGCATGTACGAGCAGCAGCCCATCTGTCTAAGGCCACGGC 291
DB 232 GACCAATACAAGGCCGAGCATGTACGAGCAGCAGCCCATCTGTCTAAGGCCACGGC 291
QY 213 AAGAAGAATGAGCAGCAGCATTAAAGGACCTGCAGCAGCGGTGGACGAGCCGAGCAG 154
DB 213 AAGAAGAATGAGCAGCAGCATTAAAGGACCTGCAGCAGCGGTGGACGAGCCGAGCAG 154
QY 292 CTGGACCAAGCAGGTGATGATCCAGCAGCAGCGCGGAGTGTGGAG 396
DB 292 CTGGACCAAGCAGGTGATGATCCAGCAGCAGCGCGGAGTGTGGAG 396
QY 153 ATGCGCCCTCAAGGGAGGCAAGAGCAGCTGCAGAAAGCTGGAAGCGCGGTGGAGGAGTG 94
DB 153 ATGCGCCCTCAAGGGAGGCAAGAGCAGCTGCAGAAAGCTGGAAGCGCGGTGGAGGAGTG 94
QY 352 GCGGTGAGCTGAGCGCCGAGCAGAGCAGCGCGGAGTGTGGAG 49
DB 352 GCGGTGAGCTGAGCGCCGAGCAGAGCAGCGCGGAGTGTGGAG 49

RESULT 15
US-09-864-761-8326/c
; Sequence 8326, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL
; FILE REFERENCE: Aomic-X-1
; CURRENT APPLICATION NUMBER: US 09/864,761
; PRIOR FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30

Search completed: April 23, 2003, 13:49:27
Job time : 110 secs

```

: PRIOR FILING DATE: 2000-02-04
: PRIOR APPLICATION NUMBER: US 60/207,456
: PRIOR FILING DATE: 2000-05-26
: PRIOR APPLICATION NUMBER: US 09/632,366
: PRIOR FILING DATE: 2000-08-03
: PRIOR APPLICATION NUMBER: GB 24263.6
: PRIOR FILING DATE: 2000-10-04
: PRIOR APPLICATION NUMBER: US 60/236,359
: PRIOR FILING DATE: 2000-09-27
: PRIOR APPLICATION NUMBER: PCT/US01/006666
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/006667
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/006664
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/006669
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/006665
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/006668
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/006663
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/006662
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/006661
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/006670
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: US 60/234,687
: PRIOR FILING DATE: 2000-09-21
: PRIOR APPLICATION NUMBER: US 09/608,408
: PRIOR FILING DATE: 2000-06-30
: PRIOR APPLICATION NUMBER: US 09/774,203
: PRIOR FILING DATE: 2001-01-29
: NUMBER OF SEQ ID NOS: 49117
: SOFTWARE: Anomax Sequence Listing Engine
: SEQ ID NO 8326
: LENGTH: 582
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: OTHER INFORMATION: MAP TO AL049829.3
: OTHER INFORMATION: EXPRESSED IN HEART, S
: OTHER INFORMATION: EXPRESSED IN BONE MAR
: OTHER INFORMATION: EXPRESSED IN LUNG, S
: OTHER INFORMATION: EXPRESSED IN PLACENTA
: OTHER INFORMATION: EXPRESSED IN BRAIN, S
: OTHER INFORMATION: EXPRESSED IN HELA, S
: OTHER INFORMATION: EXPRESSED IN ADULT LI
: OTHER INFORMATION: EXPRESSED IN FETAL LI
: IS-09-864-761-8326

```

Query Match	6.9%;	Score 41;	DB 10;	Length 582;	
Best Local Similarity	48.9%;	Pred. No. 0.0035;			
Matches 110;	Conservative	0;	Mismatches 115;	Indels 0;	Gaps
QY	172	GACATGCAAGAGCAGAGAAGTGAAGCAGGTCAGTTTCATGAGCTTCGTGTTGGCGGAGCA	231		
Db	366	GCATGATGCCACAGAGCTGAAGAAGGACGAGACACGAGCCGCCACCTGGAGCGCGATG	307		
QY	232	GACCAATACAAAGGCCGAAGCATGTACGACGACACGCCCATCTGGTCAAGGGCCACAGCGC	291		
Db	306	AAGAAGAACATGGAGCAGACCATTAAGGACCTCGAGCACCGGCTGGACGAGCCGAGCAG	247		
QY	292	CTGGACGACCGCCACTTTGACAGATCAAGCAGTACCTTTGGAGAGACGCTGCAAGAGATG	351		
Db	246	ATCGCCCTCAAGGGAGCAAGAAGCAGCTGCAGAACTCGAAGCCGGGTGCGGAGCTG	187		
QY	352	GGCGTCAAGCAGGATGTGATCCAGCACGCCGCCGGAGTGGTGGAG	396		
Db	186	GAGGTTGACCTCGAGGCCGAGCAGAGAAGCGCAACGACGACGTCTGGT	142		

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model
Run on: April 16, 2003, 12:55:35 ; Search time 38.5 Seconds
(without alignments)
4139.422 Million cell updates/sec
Title: US-09-920-953-2
Perfect score: 1067
Sequence: 1 gcgctgctgcagacgcgcg.....agccatttttcgaccacgaagcc 598

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0
arched: 908470 seqs, 133250620 residues
Total number of hits satisfying chosen parameters: 1816940
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+n2p.model -DEV=xlip
-O/cgn2_1/USPTO_spool/US09920953/runat_16042003_115019_8129/app_query.fasta_1.775
-DB=A_Geneseq_101002 -QFMT=fastan -SUFFIX=n2p.rag -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORW=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09920953 -CGN_1_114/runat_16042003_115019_8129 -NCPU=6 -ICPU=3
-NO_XLPAY -NO_MMAL -LARGEOUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV.TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : A_Geneseq_101002.*
1: /SID22/qcgdata/geneseq/geneseq-embl/AA1980.DAT.*
2: /SID22/qcgdata/geneseq/geneseq-embl/AA1981.DAT.*
3: /SID22/qcgdata/geneseq/geneseq-embl/AA1982.DAT.*
4: /SID22/qcgdata/geneseq/geneseq-embl/AA1983.DAT.*
5: /SID22/qcgdata/geneseq/geneseq-embl/AA1984.DAT.*
6: /SID22/qcgdata/geneseq/geneseq-embl/AA1985.DAT.*
7: /SID22/qcgdata/geneseq/geneseq-embl/AA1986.DAT.*
8: /SID22/qcgdata/geneseq/geneseq-embl/AA1987.DAT.*
9: /SID22/qcgdata/geneseq/geneseq-embl/AA1988.DAT.*
10: /SID22/qcgdata/geneseq/geneseq-embl/AA1989.DAT.*
11: /SID22/qcgdata/geneseq/geneseq-embl/AA1990.DAT.*
12: /SID22/qcgdata/geneseq/geneseq-embl/AA1991.DAT.*
13: /SID22/qcgdata/geneseq/geneseq-embl/AA1992.DAT.*
14: /SID22/qcgdata/geneseq/geneseq-embl/AA1993.DAT.*
15: /SID22/qcgdata/geneseq/geneseq-embl/AA1994.DAT.*
16: /SID22/qcgdata/geneseq/geneseq-embl/AA1995.DAT.*
17: /SID22/qcgdata/geneseq/geneseq-embl/AA1996.DAT.*
18: /SID22/qcgdata/geneseq/geneseq-embl/AA1997.DAT.*
19: /SID22/qcgdata/geneseq/geneseq-embl/AA1998.DAT.*
20: /SID22/qcgdata/geneseq/geneseq-embl/AA1999.DAT.*
21: /SID22/qcgdata/geneseq/geneseq-embl/AA2000.DAT.*
22: /SID22/qcgdata/geneseq/geneseq-embl/AA2001.DAT.*
23: /SID22/qcgdata/geneseq/geneseq-embl/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	118	11.1	129	21	AAG12569	Zea mays protein f
2	117	11.0	257	22	AAB95825	Human protein sequ
3	116	10.9	290	22	AU54642	Propionibacterium
4	114.5	10.2	599	22	ABG03530	Novel human diagno
5	111.5	9.9	129	21	AAG12569	Zea mays protein f
6	111.5	10.4	387	18	AAW18663	Fragmented human N
7	111	9.9	572	20	AAU04954	Mycobacterium spec
8	110.5	9.8	572	22	AAU32685	Novel human secret
9	106.5	10.0	317	22	AAU40791	Human polypeptide
10	106.5	10.0	317	22	AAU40792	Human polypeptide
11	105	9.8	242	22	AAU62666	Propionibacterium
12	105	9.8	264	22	ABG05131	Novel human diagno
13	104.5	9.3	900	22	ABG03533	Novel human diagno
14	104	9.7	305	22	AAU48541	Propionibacterium
15	103.5	9.7	1307	22	ABG65464	Drosophila melanog
16	103	9.7	280	22	ABG08082	Novel human diagno
17	102.5	9.6	229	21	AAU84582	Amino acid sequenc
18	102.5	9.6	536	21	AAU84592	Amino acid sequenc
19	102	9.6	201	22	AAU51115	Propionibacterium
20	102	9.6	539	22	ABG19415	Novel human diagno
21	101	9.0	536	21	AAU84592	Amino acid sequenc
22	100.5	9.0	264	22	ABG05131	Novel human diagno
23	100.5	9.0	292	22	AAU42102	Human polypeptide
24	100	8.9	853	22	ABG12054	Novel human diagno
25	99.5	8.9	174	22	ABG20079	Novel human diagno
26	99.5	8.9	318	22	ABG19764	Novel human diagno
27	99	9.3	243	22	AAU46489	Propionibacterium
28	99	9.3	293	22	ABG44103	Novel human diagno
29	99	9.3	330	22	ABG39598	Drosophila melanog
30	99	8.8	468	22	AAU46227	Propionibacterium
31	99	8.8	1061	20	AAU87504	Human N-methyl-D-a
32	98.5	8.8	310	22	AAO00093	Human polypeptide
33	98.5	8.8	387	18	AAW18663	Fragmented human N
34	98.5	9.2	513	21	AAU24443	Mycobacterium tube
35	98.5	8.8	1212	22	AAU87503	Human N-methyl-D-a
36	98	8.7	175	22	ABU11118	Human secreted pro
37	98	9.2	229	22	AAU47602	Propionibacterium
38	98	9.2	241	20	AAU73994	Human prostate tum
39	98	8.7	332	21	AAU82703	Tick derived cyste
40	97.5	9.1	640	22	ABG03549	Novel human diagno
41	97	8.6	195	22	AAU30711	Novel human secret
42	97	8.6	346	22	AAU19442	Human diagnostic a
43	96.5	9.0	208	22	ABG28202	Novel human diagno
44	96.5	9.0	270	22	AAU56363	Propionibacterium
45	96.5	8.6	336	23	ABP41746	Human ovarian anti

ALIGNMENTS

RESULT 1
AAG12569
ID AAG12569 standard; Protein; 129 AA.
XX
AC AAG12569;
XX
DT 17-OCT-2000 (first entry)
XX
DE Zea mays protein fragment SEQ ID NO: 11732.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence; corn.
XX
OS Zea mays subsp. mays.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.

PR 24-SEP-1999; 99US-0155659.
 PR 28-SEP-1999; 99US-0156458.
 PR 29-SEP-1999; 99US-0156596.
 PR 04-OCT-1999; 99US-0157117.
 PR 05-OCT-1999; 99US-0157753.
 PR 06-OCT-1999; 99US-0157865.
 PR 07-OCT-1999; 99US-0158029.
 PR 08-OCT-1999; 99US-0158232.
 PR 12-OCT-1999; 99US-0158369.
 PR 13-OCT-1999; 99US-0159293.
 PR 13-OCT-1999; 99US-0159294.
 PR 13-OCT-1999; 99US-0159295.
 PR 14-OCT-1999; 99US-0159329.
 PR 14-OCT-1999; 99US-0159330.
 PR 14-OCT-1999; 99US-0159331.
 PR 14-OCT-1999; 99US-0159637.
 PR 18-OCT-1999; 99US-0159638.
 PR 18-OCT-1999; 99US-0159584.
 PR 21-OCT-1999; 99US-0160741.
 PR 21-OCT-1999; 99US-0160767.
 PR 21-OCT-1999; 99US-0160768.
 PR 21-OCT-1999; 99US-0160770.
 PR 21-OCT-1999; 99US-0160814.
 PR 21-OCT-1999; 99US-0160815.
 PR 22-OCT-1999; 99US-0160980.
 PR 22-OCT-1999; 99US-0160981.
 PR 22-OCT-1999; 99US-0160989.
 PR 25-OCT-1999; 99US-0161404.
 PR 25-OCT-1999; 99US-0161405.
 PR 25-OCT-1999; 99US-0161406.
 PR 26-OCT-1999; 99US-0161359.
 PR 26-OCT-1999; 99US-0161360.
 PR 26-OCT-1999; 99US-0161361.
 PR 28-OCT-1999; 99US-0161920.
 PR 28-OCT-1999; 99US-0161992.
 PR 28-OCT-1999; 99US-0161993.
 PR 28-OCT-1999; 99US-0162142.
 PR 28-OCT-1999; 99US-0162142.

Alignment Scores:
 Pred. No.: 0.00838
 Score: 118.00
 Percent Similarity: 45.93%
 Best Local Similarity: 31.85%
 Query Match: 11.08%
 DB: 21
 Length: 129
 Matches: 43
 Conservative: 19
 Mismatches: 37
 Indels: 36
 Gaps: 11

US-09-920-953-2 (1-598) x AAG12569 (1-129)

QY 80 TGGCGCGCCAGACGATGAAGCTGGCGG-----TTGACACCTTCT 121
 D 1 TTPAlaAlaArg-----TTPArgHisCysAlaGlyArgProThrAlaProThr 16
 QY 122 ACATGAAGCTGCTGGCTGACC-----CGAGCTGCTGC----- 154
 DB 17 SerThrArgCysTrpAlaSerIleCysSerAlaArgProCysCysAlaSerArgAla 36
 QY 155 -----CCTCTTCAGTCC-----TGGACATGC---AAGACGAGA 187
 DB 37 PheArgTrpProSerSerThrProSerArgSerArgSerTrpCysCysTrpGlySerArg 56
 QY 188 ACATGACGACGTCAGTTCATGCTTCTGCTTGGCGGACGACGACCAATACAGGGCC 247
 DB 57 ArgArgSerProAlaProArgProSerSerCysThrAla-----SerSerArgArgSer 74
 QY 248 GAAGCATGTACACGACACGCCATCTGCTCAAGGGCCACGCGCTGGACCCGCGCACT 307
 DB 75 SerArgThrProSerThrSerPro-----SerArgSerSerCysArgArgArgAlaSer 92
 QY 308 TTGACAGATCA---AGCAGTACCTTGGAGACGCTGCAAGAGATGGCGCTCAAGCAGG 364
 DB 93 CysSerArgAlaProThrSerProProSerArgTrpCysCysThrTrpTrpSerAlaGly 112
 QY 365 ATGTGATCCACGACGCGCGCGGAGTGG-----TGGAGTCCACCC 403

Db 113 Ser---SerSerThrArgSerAlaTrpAlaCysTrpGlyProPro 126
 RESULT 2
 AAB95825
 ID AAB95825 standard; Protein; 257 AA.
 XX
 AC AAB95825;
 XX
 DT 26-JUN-2001 (first entry)
 XX
 DE Human protein sequence SEQ ID NO:18834.
 XX
 KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.
 XX
 OS Homo sapiens.
 XX
 PN EPI074617-A2.
 XX
 PD 07-FEB-2001.
 XX
 PF 28-JUL-2000; 2000EP-0116126.
 XX
 PR 29-JUL-1999; 99JP-0248036.
 PR 27-AUG-1999; 99JP-0300253.
 PR 11-JAN-2000; 2000JP-0118776.
 PR 02-MAY-2000; 2000JP-0183767.
 PR 09-JUN-2000; 2000JP-0241899.
 XX
 PA (HELI-) HELIX RES INST.
 XX
 PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 XX
 DR WPI; 2001-318749/34.
 XX
 PT Primer sets for synthesizing polynucleotides, particularly the 5602
 PT full-length cDNAs defined in the specification, and for the detection
 PT and/or diagnosis of the abnormality of the proteins encoded by the
 PT full-length cDNAs -
 PS Claim 8; SEQ ID 18834; 2537pp + CD ROM; English.
 CC The present invention describes primer sets for synthesizing 5602
 CC full-length cDNAs defined in the specification. Where a primer set
 CC comprises: (a) an oligo-dr primer and an oligonucleotide complementary
 CC to the complementary strand of a polynucleotide which comprises one of
 CC the 5602 nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in
 CC the specification. The primer sets can be used in antisense therapy and
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
 CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
 CC represent oligonucleotides, all of which are used in the exemplification
 CC of the present invention.
 XX
 SQ Sequence 257 AA;

Alignment Scores:
 Pred. No.: 0.00139
 Score: 117.00
 Percent Similarity: 38.10%
 Best Local Similarity: 28.57%
 Length: 257
 Matches: 60
 Conservative: 20
 Mismatches: 72

CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

Sequence 599 AA;

Alignment Scores:		
Pred. No.:	0.00352	Length:
Score:	14.50	Matches:
Percent Similarity:	33.18%	Conservative:
Best Local Similarity:	26.73%	Mismatches:
Query Match:	10.20%	Indels:
DB:	22	Gaps:
		12
		599

US-09-920-953-2 (1-598) x ABG03530 (1-599)

Qy	583	TGGCTGATGGCTCTTCAATGCAGCAAGGGCTGG-----CAAACGATTTGCAGGCTC	533
Db	74	TrpThrProHisCysProAlaSerAlaAlaTrpGlyAlaGluGlnProValArgSer	93
Qy	532	CTGGGCCCCCTGGGCGGCTC-----GGCAACTGGCGCTCCCGCTCAAAATC	485
Db	94	TrpGlyProArgAlaSerGlnSerHisCysProGlyGlyLeuArgAlaProProGly	113
Qy	484	GAT---GGAATGACGCGCTCAGGCTGGGTTTGAATGAAATCACTTGGTGGCGCAGT	428
Db	114	SerValArgCysSerThrGln**AspCysSerSerValArgProAlaTrpSerArgSer	133
Qy	427	-----TGTTGGNNAAGTCAAAATTCGTCCGCGGGTGG	398
Db	134	**GlyAlaCys***GlnVal***ProArgCysProCysArgThrProAlaThrGlyTrp	153
Qy	397	ACTCCACCACTCCGGCGGCTGTGGA-----TCACATCTCGCT	359
Db	154	AlaProProGlnGlyArgCysGlyProThrThrAlaProGlySerThrGlyProAla	173
Qy	358	TGAGCGCCATCTCTTGCGAGGCTCTCTCCAAGGTACTGCTTGATCTTGTCAAAGTGGCGGT	299
Db	174	GlyArgAlaSerLeuCysCysProArgArgAlaHisLeu-----	186
Qy	298	GGTCCAGCGGTGGCCCTTGACCAGATGGCGGTGCGCTGCATCGTCTCGGCCCT	242
Db	187	---ProGly**TrpPro-----GlnLysLeuIleCysAlaHisProGly	200
Qy	241	-----TGTATTGGTCTGCTCCGCCAACCAAGAGTCATGAACCTGCCTTCATCT	188
Db	201	AlaLysSerLeuGlyLeuAlaCysGlnProHisArgGlyLysGlyThrProIleGluGly	220
Qy	187	TCTGCTCTTCATGTCACGGACTCGAAGAGGGCAGCAGCTCCGGGTGACGAGCACCT	128
Db	221	ProAlaCysGlyThr**GlyGlyArgArgGlySerGlyCysProGlyArgProHisThr	240
Qy	127	TATCGTAGAAGGTGTCAACCGCCAGCTTCATCGCTTCTGCGCGCCCATCATCAACA	68
Db	241	---ArgArgArgCys**ProProAlaProCys-----GlyArgArgSer	254
Qy	67	GCTTCTTGGCCCCGCATCCCGCTTCGCGCTCTCTGTGTGCTGGTGGCGG	17
Db	255	AlaGlySerAlaHis---ProAlaArgPro-----TrpPro	265

polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic amino acid sequences of the invention.

Qy	293	-----TGACACCGCCACCTTTGACAAGATCAACGAGTACCTTG	331
Db	135	TrpProLeuProProCysTyrTrpProAlaTrpThrAlaSerArgThrArgSerSerArg	154
Qy	332	GAGAGACGCTGCAAGAGATGGCGCTCAACAGGATGTGATCCAGCAGCGCGCGGAGTGG	391
Db	155	ProArgArgSerThrArgThrSerThrSerSerLeuLeuArgSerMetProPro---Ttp	173
Qy	392	-----TGG-----AGTCCACCGCGCAGGAATTG	415
Db	174	IleThrCysArgAlaThrTrpValArgSerSerThrIleSerArgProThrThrSerPhe	193
Qy	416	ACTTNCCCAACAACCTGGCACCCA	439
		:::	
Db	194	SerLeuProValThrCysSerPro	201

RESULT 4
ABG03530
ID ABG03530 standard: protein: 599 AA.

ABG03530:

XX
DT 13-FEB-2002 (first entry)XX
DE Novel human diagnostic protein #3521.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder;
KW food supplement; medical imaging; diagnostic; genetic disorder;
XX

XX
OS
Homo sapiens.

XX PN WO200175067-A2.

XX
PD
11-OCT-2001XX
PF 30-MAR-2001: 2001WO-US08631.XX
PR 31-MAR-2000: 2000US-0540217

PR 23-AUG-2000; 2000US-0649167.
YY

PA (HYSE-) HYSEQ INC.
YY

PI Drmanac RT, Liu C, Tang YT;

DR WPI; 2001-639362/73.

XX

diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits and to assess biodiversity.

PS Claim 20; SEQ ID No 33889; 103pp; English.

The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. ABC00010-ABC03077 represent novel human diagnostic amino acid sequences of the invention.

OS	PN	PD	PF	XX	Zea mays subsp. mays.	99US-0140991
OS	PN	PD	PF	XX	25-FEB-2000; 2000EP-0301439.	99US-0141287.
OS	PN	PD	PF	XX	25-FEB-1999;	99US-0141287.
OS	PN	PD	PF	XX	05-MAR-1999;	99US-0141842.
OS	PN	PD	PF	XX	09-MAR-1999;	99US-0142154.
OS	PN	PD	PF	XX	23-MAR-1999;	99US-0142055.
OS	PN	PD	PF	XX	25-MAR-1999;	99US-0142390.
OS	PN	PD	PF	XX	29-MAR-1999;	99US-0142803.
OS	PN	PD	PF	XX	01-APR-1999;	99US-0142920.
OS	PN	PD	PF	XX	06-APR-1999;	99US-0142977.
OS	PN	PD	PF	XX	11-APR-1999;	99US-0143542.
OS	PN	PD	PF	XX	16-APR-1999;	99US-0143624.
OS	PN	PD	PF	XX	18-APR-1999;	99US-0144005.
OS	PN	PD	PF	XX	19-APR-1999;	99US-0144085.
OS	PN	PD	PF	XX	21-APR-1999;	99US-0144086.
OS	PN	PD	PF	XX	23-APR-1999;	99US-0144325.
OS	PN	PD	PF	XX	28-APR-1999;	99US-0144331.
OS	PN	PD	PF	XX	30-APR-1999;	99US-0144332.
OS	PN	PD	PF	XX	04-MAY-1999;	99US-0144333.
OS	PN	PD	PF	XX	05-MAY-1999;	99US-0144335.
OS	PN	PD	PF	XX	06-MAY-1999;	99US-0144352.
OS	PN	PD	PF	XX	06-MAY-1999;	99US-0144632.
OS	PN	PD	PF	XX	07-MAY-1999;	99US-0144884.
OS	PN	PD	PF	XX	11-MAY-1999;	99US-0144814.
OS	PN	PD	PF	XX	14-MAY-1999;	99US-0145086.
OS	PN	PD	PF	XX	14-MAY-1999;	99US-0145088.
OS	PN	PD	PF	XX	14-MAY-1999;	99US-0145085.
OS	PN	PD	PF	XX	14-MAY-1999;	99US-0145087.
OS	PN	PD	PF	XX	14-MAY-1999;	99US-0145089.
OS	PN	PD	PF	XX	14-MAY-1999;	99US-0145192.
OS	PN	PD	PF	XX	18-MAY-1999;	99US-0145118.
OS	PN	PD	PF	XX	19-MAY-1999;	99US-0145218.
OS	PN	PD	PF	XX	21-MAY-1999;	99US-0145214.
OS	PN	PD	PF	XX	21-MAY-1999;	99US-0145224.
OS	PN	PD	PF	XX	24-MAY-1999;	99US-0145276.
OS	PN	PD	PF	XX	25-MAY-1999;	99US-0145913.
OS	PN	PD	PF	XX	27-MAY-1999;	99US-0145918.
OS	PN	PD	PF	XX	28-MAY-1999;	99US-0145919.
OS	PN	PD	PF	XX	01-JUN-1999;	99US-0145951.
OS	PN	PD	PF	XX	03-JUN-1999;	99US-0146386.
OS	PN	PD	PF	XX	04-JUN-1999;	99US-0146388.
OS	PN	PD	PF	XX	07-JUN-1999;	99US-0146389.
OS	PN	PD	PF	XX	08-JUN-1999;	99US-0147038.
OS	PN	PD	PF	XX	10-JUN-1999;	99US-0147038.
OS	PN	PD	PF	XX	10-JUN-1999;	99US-0147204.
OS	PN	PD	PF	XX	14-JUN-1999;	99US-0147302.
OS	PN	PD	PF	XX	16-JUN-1999;	99US-0147192.
OS	PN	PD	PF	XX	16-JUN-1999;	99US-0147260.
OS	PN	PD	PF	XX	17-JUN-1999;	99US-0147303.
OS	PN	PD	PF	XX	17-JUN-1999;	99US-0147303.
OS	PN	PD	PF	XX	18-JUN-1999;	99US-0147416.
OS	PN	PD	PF	XX	18-JUN-1999;	99US-0147493.
OS	PN	PD	PF	XX	18-JUN-1999;	99US-0147935.
OS	PN	PD	PF	XX	18-JUN-1999;	99US-0148171.
OS	PN	PD	PF	XX	18-JUN-1999;	99US-0148319.
OS	PN	PD	PF	XX	18-JUN-1999;	99US-0148341.
OS	PN	PD	PF	XX	18-JUN-1999;	99US-0148565.
OS	PN	PD	PF	XX	18-JUN-1999;	99US-0148684.
OS	PN	PD	PF	XX	18-JUN-1999;	99US-0149175.
OS	PN	PD	PF	XX	18-JUN-1999;	99US-0149368.
OS	PN	PD	PF	XX	18-JUN-1999;	99US-0149426.
OS	PN	PD	PF	XX		

PR 13-SEP-1999: 99US-0153758.
 PR 15-SEP-1999: 99US-0154018.
 PR 16-SEP-1999: 99US-0154039.
 PR 20-SEP-1999: 99US-0154779.
 PR 22-SEP-1999: 99US-0155139.
 PR 23-SEP-1999: 99US-0155486.
 PR 24-SEP-1999: 99US-0156659.
 PR 28-SEP-1999: 99US-0156458.
 PR 29-SEP-1999: 99US-0156596.
 PR 04-OCT-1999: 99US-0157117.
 PR 05-OCT-1999: 99US-0157753.
 PR 06-OCT-1999: 99US-0157865.
 PR 07-OCT-1999: 99US-0158029.
 PR 08-OCT-1999: 99US-0158232.
 PR 12-OCT-1999: 99US-0158369.
 PR 13-OCT-1999: 99US-0159293.
 PR 13-OCT-1999: 99US-0159294.
 PR 14-OCT-1999: 99US-0159329.
 PR 14-OCT-1999: 99US-0159330.
 PR 14-OCT-1999: 99US-0159331.
 PR 14-OCT-1999: 99US-0159637.
 PR 18-OCT-1999: 99US-0159638.
 PR 21-OCT-1999: 99US-0159584.
 PR 21-OCT-1999: 99US-0160741.
 PR 21-OCT-1999: 99US-0160767.
 PR 21-OCT-1999: 99US-0160768.
 PR 21-OCT-1999: 99US-0160770.
 PR 21-OCT-1999: 99US-0160814.
 PR 21-OCT-1999: 99US-0160815.
 PR 22-OCT-1999: 99US-0160980.
 PR 22-OCT-1999: 99US-0160981.
 PR 22-OCT-1999: 99US-0160989.
 PR 23-OCT-1999: 99US-0161404.
 PR 23-OCT-1999: 99US-0161405.
 PR 25-OCT-1999: 99US-0161406.
 PR 26-OCT-1999: 99US-0161359.
 PR 26-OCT-1999: 99US-0161360.
 PR 26-OCT-1999: 99US-0161361.
 PR 28-OCT-1999: 99US-0161920.
 PR 28-OCT-1999: 99US-0161992.
 PR 28-OCT-1999: 99US-0161993.
 PR 29-OCT-1999: 99US-0162142.

Alignment Scores:

Pred. No.: 0.00412 Length: 129
 Score: 111.50 Matches: 40
 Percent Similarity: 38.85% Conservative: 14
 Best Local Similarity: 28.78% Mismatches: 39
 Every Match: 9.94% Indels: 46
 Gaps: 21

US-09-920-953-2 (1-598) x AAG12569 (1-129)

QY 410 TCGTCGGGGTGCACCTCCCGCGGCGTGGATCATCTCTGACGCC 351
 Db 9 AlaGlyArgProThrAlaProThrSerThrArgCysTrpAlaSerIleCys----- 25
 QY 350 ATCTCTGACGCTCTCCAGGTACTGCTTCATCTGTCAAAG---TGGCGGTGGTCC 294
 Db 26 -----SerAlaArgProCysCysAlaSerArgAlaPheArgTrp-Pr 40
 QY 293 AGCCGTGGCCCTTGACCATGGCGTGGCTGCTACATGCTTCGGCCCTTGATGG 234
 Db 40 oSerSerThrPro-----SerArgSerArgSerTrpCys----- 51
 QY 233 TCTGCTCCGCCAAACACAGACTCATGCACTTCACCTGCTTCATCTCTCTGTCATG 174
 Db 52 -----Cy 52
 QY 173 TCCAGGACTCTGAAGAGGCGACAGCTCGGGTCAAGCCA-----GCACCTTA 126
 Db 52 sTrpGlySerArgArgSerProAlaPro---ArgProSerSerCysThrAlaSer 71

QY 125 TCGTAGAAGGTGTCACCGCCAGCTTCATGCTTCTGCGCGCCAGGTCATCAACACGC 66
 Db 71 rArgArgSerSerArgThrProSerProSerArgSerSerCysArgArgAl 91
 QY 65 TTCTTGGCGCGCATCCCGCTT---CGCGCGTCTCTGTGTGTC-----TGGTGG 20
 Db 91 aserCysSerArgAlaProThrSerProSerArgTrpCysCysThrTrpTrp 109
 RESULT 6
 AAW18663
 ID AAW18663 standard; Protein; 387 AA.
 XX
 AC AAW18663;
 XX
 DT 24-JUL-1997 (first entry)
 XX
 DE Fragmented human NF-H gene +2 frameshift mutant product.
 XX
 KW Frameshift mutation product; GAGA motif; somatic mutation; diagnosis;
 KW detection; antibody; probe; cancer; neoplasia; neurodegenerative;
 KW Parkinson's; Alzheimer's disease; Pick's; Huntington's disease;
 KW Down's syndrome; frontal lobe dementia; progressive supranuclear palsy;
 KW PSP; amyotrophic lateral sclerosis; multiple sclerosis; MS;
 KW cardiovascular; rheumatoid arthritis.
 XX
 OS Homo sapiens.
 XX
 FH Key
 FT Location/Qualifiers
 FT Misc-difference 1..387
 FT /note= "X corresponds to a stop codon in the
 FT accompanying DNA file, AAT69796"
 XX
 PN W09712992-A2.
 XX
 PD 10-APR-1997.
 XX
 PF 02-OCT-1996; 96WO-IB01106.
 XX
 PR 11-JAN-1996; 96US-0009832.
 PR 02-OCT-1995; 95GB-0020080.
 XX
 PA (ROYA-) ROYAL NETHERLANDS ACAD ARTS & SCI.
 PA (UYUO-) UNIV ROTTERDAM ERASMUS.
 PA (UYUT-) UNIV STATE UTRECHT.
 XX
 PI Burbach JPH, Grosveid FG, Van Leeuwen FW;
 XX
 WPI: 1997-226235/20.
 DR N-PSDB; AAT69795.
 DR
 XX
 PT Use of mutant genes having frame:shift mutation(s) - for developing
 PT prods. for the diagnosis, prevention and treatment of associated
 PT diseases, e.g. cancer or neurodegenerative disease
 XX
 PS Claim 22; Fig 9; 123pp; English.
 XX
 CC AAW18663 and AAW18664 are +2 and +1 frameshift mutations, respectively,
 CC of a sequence comprising fragments of the coding sequence of the
 CC human neurofilament subunit NF-H gene corresponding to nucleotides
 CC 1-1162 of the wild-type NF-H gene. This region contains GAGAG motifs.
 CC Frameshift mutants of the tau, ubiquitin, apolipoprotein E,
 CC microtubule-associated protein 2 (MAP-2), neurofilament subunit L, M
 CC and H and amyloid A4 genes are claimed. All these genes share a common
 CC GAGAN motif (N= A, G, C or T), which is the site of common GA
 CC dinucleotide deletion(s) that cause neurodegenerative disorders.
 CC Antigenic peptides used for the production of antibodies, and small
 CC nucleic acid sequences derived from frameshift mutants are used in the
 CC diagnosis, prevention and treatment of cancer and neurodegenerative
 CC diseases, e.g. Parkinson's disease, Alzheimer's disease, Down's
 CC syndrome, frontal lobe dementia (Pick's disease), progressive
 CC supranuclear palsy (PSP), amyotrophic lateral sclerosis, Huntington's
 CC disease, multiple sclerosis, and other degenerative diseases such as

CC cardiovascular disease and rheumatoid arthritis.

XX Sequence 387 AA;

Alignment Scores:

Pred. No.: 0.00622 Length: 387
Score: 111.50 Matches: 40
Percent Similarity: 35.95% Conservative: 15
Best Local Similarity: 26.14% Mismatches: 49
Query Match: 10.45% Indels: 49
DB: 18 Gaps: 7

US-09-920-953-2 (1-598) x AAH18663 (1-387)

Qy 11 CAGACGCGCCACACACACAGACGCGGAGCGGATGCGGCGGCAAGAGCTGT 70

Db 161 GlnAlaProGlnAlaProThrArgTrpArg***AlaThrGlyArgAla-- 179

Qy 71 TTGATGACCTGGCGCGCCAGAGCATGAAGCTGGCGGTGCACACCTCTACGATAAGG 130

180 -----AlaTrp-----TTPArgTrpProHisAlaValArg 190

Qy 131 TGCTGGCTGACCGGAGCTGCTGC-----CCTTCTCGAGTCCC----- 169

Db 191 -----ArgSerCysArgArg***ThrThrAlaSerProGlyThrSerThr 206

Qy 170 -----TGACATGCAAGCAGCAGAGATGAAGC 196

Db 207 ArgCysGlySerTrpArgArgThrThrAlaAlaTrpArgAlaArgLeuArgCysGly 226

Qy 197 AGGTCAAGTTCATGAGCTGTGTTGGCGGAGCAGCAATACAAAGGCGGAGCATGT 256

Db 227 SerSerArgAlaAlaProLeuTrpAlaSerCysThrSerAlaArgSerAlaArgCys 246

Qy 257 ACGAGCACACGCCATCTGTGTCAGG-----GCCACGGCTGCACACGCCCACT 307

Db 247 AlaAlaArgCysCysAlaTrpAlaArgAlaValSerTyrAlaTrpSerArgSerThr 266

Qy 308 TTGACAAGTCAAGCAGTACCTTG-----TTPArgTrpProGlySerGluArg 286

Db 267 CysSerArgThrSerArgThrCysAlaSerAla***ThrThrArgProGlySerGluArg 286

Qy 332 -----GAGACACGCTGCAAGAGATGGCGGCTCAAGCAGG 364

Db 287 ArgProArgArgProAlaArgTrpArgAlaSerArg 299

RESULT 7

AY04954

AAI04954 standard; Protein; 572 AA.

AC AAY04954;

XX 06-JUL-1999 (first entry)

DE Mycobacterium species protein sequence 41T#2.

XX Secreted protein; Mycobacterium; primer; PCR; amplification; probe;

XX hybridisation; detection; vaccine; immunisation; infection.

XX Mycobacterium sp.

OS WO9909186-A2.

XX 25-FEB-1999.

XX 14-AUG-1998; 98WO-FR01813.

XX 11-SEP-1997; 97FR-0011325.

PR 14-AUG-1997; 97FR-0010404.

XX (INSP) INST PASTEUR.

XX Gicquel B, Lim EM, Pelicic V, Portnoi D, Goguet de la Salmoniere Y;

PI Guigueno A;

XX WPI; 1999-181045/15.

DR N-PSDB; AAX34206.

XX Mycobacterial DNA vectors containing reporter constructs - for

PT identifying coding or promoter sequences involved in

PT infection-associated protein expression

XX Claim 32; Fig 41T; 309pp; French.

XX Sequences AAY04742-Y05000 and AAY07201-Y07204 represent secreted

CC encoding nucleotide sequences can be used as primers and probes for

CC methods for detecting and identifying mycobacteria, especially belonging

CC to the M. tuberculosis complex. The encoded proteins can be used in

CC vaccines for immunisation against a bacterial or viral infection.

XX SQ Sequence 572 AA;

Alignment Scores:

Pred. No.: 0.00814 Length: 572
Score: 111.00 Matches: 58
Percent Similarity: 30.25% Conservative: 14
Best Local Similarity: 24.37% Mismatches: 87
Query Match: 9.89% Indels: 79
DB: 20 Gaps: 12

US-09-920-953-2 (1-598) x AAY04954 (1-572)

Qy 533 CTGGGCGCCCTGGGCGGCTCGGCACTGGCGCTCCCGCTCAAAATCGATGGAATGAG 474

Db 313 ProValProPro-----TTPArgTrpProGlySerValProGlyProAla 319

Qy 473 CGCTCAGCTGGGTTGGTTAATCAAAATCAGTTGGTGGCGCAGTTGTTGGGNAAGTCA 414

Db 320 AspTrpArgTrpGlyGluLeuAlaGlySer***SerArgArgSerValProGlyProAla 339

Qy 413 AATTCGTCGCGGGTGGACTCCACCACCTCCGCGGCGTCTGCATCACATCTGCTGTGAGC 354

Db 340 AspCysArgProValAlaGlyArgGlyAlaAlaProCysTrpArgSerSerThrAlaThr 359

Qy 353 ---CCCATCTCTTCAGCGTC-----TCTCAAGTACTGC----- 321

Db 360 ValProProSerCysSerProGlyArgAlaProAlaCysCysAspArgValGlnThrPro 379

Qy 320 -----TTGATCTTGTCAAGTGG-----CGGTGCTCAGCCGCGG 285

Db 380 ***HisArgProProLleSerValProThrSerTrpCysGlyProTrp***ThrProAla 399

Qy 284 CCCTTGACCATGGCGTGTGCGTGTGATGCTTGGCGCTTGTATTGG-----TCT 231

Db 400 ProArgThr***TrpCysCysPrometAlaMetTrpProProLysAsnTrpTrpProGly 419

Qy 230 GCTCCG-----TGTCCAGGACTCGAAGAGGCGCAGCTCCGGGTCA 138

Db 420 ValProArgArgSerAlaGlyAlaSerThrTrpTrpProCysArgProAspArgTrpCys 439

Qy 224 -----CCAAACACGAAGCTCATGAACTTACCTTCATCTTCTGCTCTTGA-- 176

Db 440 ArgGlyTrpProArgTrpProCysMet-ThrArgProAlaArgProSerThrThrAlaTh 459

Qy 175 -----TGTCCAGGACTCGAAGAGGCGCAGCTCCGGGTCA 138

Db 459 rAlaTrpProValProProValLeuProGlyThrAspArgCysAlaLeuProProLysArg 479

Qy 137 G-----CCAGCACCTTATCGTAGAAGGTGTCAACCGCAGCTTCATCCCTTCT 90

Db 479 gArg***ProGlyProValProAlaSerArg-----AlaThrValTrpVa 494

Qy 89 CGCGCGCCAGGTCAACACAGCTTCTTGGCGCGCATCCCGCTTCCCGCTCTCTGTG 30

Db 494 lSerArgAlaThrArgCys***SerSerProThrMetSerProArgProSerValTr 514

PT such as central nervous system injuries -
PS Example 2; SEQ ID NO 5723; 10078pp; English.
XX
CC The invention relates to human nucleic acids (AAI57798-AAI61369) and
CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
X
J Sequence 317 AA;

Alignment Scores:

Pred. No.:	0.0196	Length:	317
Score:	106.50	Matches:	64
Percent Similarity:	33.47%	Conservative:	15
Best Local Similarity:	27.12%	Mismatches:	68
Query Match:	9.98%	Indels:	89
DB:	22	Gaps:	12

US-09-920-953-2 (1-598) x AAM40792 (1-317)	
QY 6 GCCTGCAGCGCGCCACACAGCACCACAGAGCGCGGAGCGGGATGCGG-----	56
DB 62 AleAlaGlnSerProHisArgArgGlyArgHisGlyGlyGlyAlaGlyLeuProPro 81	
QY 57 -----GCGCAAGAGCTGTTGTGACCT-----	80
DB 82 ProArgSerProArgPheProGlnGlnSerVal-----ProAlaSerThrSerThra 99	
QY 81 -----GGCGCG 86	
DB 100 ArgGlyProArgValSerArgArgLeuProGlnHisProGlyProArgGlyArg 119	
QY 87 CGGAGAGGATGAGCGGTGGTGGTACACCTTCTACGATAAGGTGCTGCTGACCGCGA 146	
DB 120 ArgArgArg-----ProGly 124	
QY 147 GCTGCTCCCTCTTCCAGTCCCTGGACATGCAAGACGACAGATGAAGCAGGTCAAGTT 206	
DB 125 AlaGly-----ValGlyAlaProArgArgGlyArgAlaGlyGlnAlaGlyLeuLeu 142	
QY 207 CATGAGCTTCTGCTTTGGCGGACGACCAATACAGGGCGGACGATGTACGACGCACA 266	
DB 143 -----GlyArgGlnGlyGlnGlyGlyArgGlyAlaGluArgGlu 155	
QY 267 CGGCCATCTCGTCAA-----GGCCACGCGCTGGACCGCCACTTTCACAGAT 317	
DB 156 ArgAlaAlaLeuGlnAlaArgArgGlyArgProGlyProGluPro-----Asp 172	
QY 318 CAAGCAGTA-----CCTTGGAGAGCGCTGCAAGAGATGGCGCTCAAGCAGGATGT 368	
DB 173 GlnSerCysGlyArgProArgArgAlaAlaAlaAlaProGlyArg-----AlaProAla 191	
QY 369 GATCCACGACCGCGCGCGTGGAGTCCACCGCGGACGAATTTGACTTNCACCAACAA 428	
DB 192 AspPro--GlnProAlaProArgProAlaProAlaProAlaProAlaProAla 211	
QY 429 CTCGG-----CACCCAACTGATTTTCATTAAACCCACCGCGCTGAGC-----	472
DB 211 spAlaProAlaProAlaProAlaProAlaProProProProHisLeuGlyAlaLeuT 231	

QY 473 -----GCTCATTCATCATGATTTTCAGC 494	
DB 231 hrAlaGlySerGlyGluArgGlnSerGlnProArgAlaGluThrLeuArgLeuGlyA 251	
QY 495 GGGAGCGCCAGTTCCGAGCGCGCCAGCGGCGCCAGGAGCGCT 538	
DB 251 rgGlyAlaProLeuProArgAlaGluArgGlyGlyArgPro 265	
RESULT 11	
AAU62666	
ID AAU62666 standard; Protein; 242 AA.	
XX	
AC AAU62666;	
XX	
DT 27-FEB-2002 (first entry)	
XX	
DE Propionibacterium acnes immunogenic protein #23562.	
XX	
KW SAPHO syndrome; synovitis; acne; pustulosis; hypertonosis; osteomyelitis;	
KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;	
KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;	
KW dermatological; osteopathic; neuroprotectant.	
XX	
OS Propionibacterium acnes.	
XX	
PN HQ200181581-A2.	
XX	
PD 01-NOV-2001.	
XX	
PF 20-APR-2001; 2001WO-US12865.	
XX	
PR 21-APR-2000; 2000US-199047P.	
PR 02-JUN-2000; 2000US-208841P.	
PR 07-JUL-2000; 2000US-216747P.	
XX	
PA (CORI-) CORIXA CORP.	
XX	
PI Skelky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;	
PI L'maisonneuve J, Zhang Y, Jen S, Carter D;	
XX	
DR N-PSDB; AAS59628.	
XX	
PT Propionibacterium acnes polypeptides and nucleic acids useful for	
PT vaccinating against and diagnosing infections, especially useful for	
PT treating acne vulgaris	
XX	
PS Example 1; SEQ ID No 23861; 1069pp; English.	
XX	
CC Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic	
CC polypeptides. The proteins and their associated DNA sequences are used in	
CC the treatment, prevention and diagnosis of medical conditions caused by	
CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,	
CC pustulosis, hypertonosis and osteomyelitis), uveitis and endophthalmitis.	
CC P. acnes is also involved in infections of bone, joints and the central	
CC nervous system, however it is particularly involved in the inflammatory	
CC lesions associated with acne vulgaris. A method for detecting the	
CC presence or absence of P. acnes in a patient comprises contacting a	
CC sample with a binding agent that binds to the proteins of the invention	
CC and determining the amount of bound protein in the sample. The	
CC polypeptides may be used as antigens in the production of antibodies	
CC specific for P. acnes proteins. These antibodies can be used to	
CC downregulate expression and activity of P. acnes polypeptides and	
CC therefore treat P. acnes infections. The antibodies may also be used as	
CC diagnostic agents for determining P. acnes presence, for example, by	
CC enzyme linked immunosorbent assay (ELISA).	
CC Note: The sequence data for this patent did not form part of the printed	
CC specification, but was obtained in electronic format directly from WIPO	
CC at ftp.wipo.int/pub/published_pct_sequences.	
XX	
SQ Sequence 242 AA;	

Alignment Scores:

Pred. No.: 0.0256 Length: 242
Score: 105.00 Matches: 59
Percent Similarity: 37.3% Conservative: 15
Best Local Similarity: 29.80% Mismatches: 68
Query Match: 9.84% Indels: 57
DB: 22 Gaps: 11

US-09-920-953-2 (1-598) x AU62666 (1-242)

QY 62 AGAAGCTGTTGATGACCTGGCGCGCAGAGGATGAAGCTGGCGTTGACACCTTCT 121
DB 1 ArgSerCysProSerArgTrpGluSerProSerAla-----TrpArgProThrProThr 18
QY 122 AGATAAGTGTGGCTGACCGGAGCTGCTCCCTTCTCGAGTCCCTGGACA----- 175
DB 19 Profile-----SerSerCysAlaTrpThrProAla 28
QY 176 -----TGCAAGAGCAGAGATGAGCAGCTCAAGTTCATGAGCTTCGTGT 220
DB 29 ArgGlyAlaAlaSerTrpAspArgLeuArgIleSerSerThrSerSerArgThrSerAla 48
QY 221 TTGGCGGAGCAGACCAATAAAGG---GCCGAGCATGTACGACGACGACGACCTCTGG 277
DB 49 ThrAlaGlyProProArgSerArgProAlaProGlyThrThrSerThrProValThr 68
QY 278 TCAAGGCGCAGCGCTGGACCGCCACTTTGACAAGATCAAGAGTACCTTGGAGAGA 337
DB 69 SerArgCys-----ArgSerThrProSerThrProArgSerThrThrProArg 84
QY 338 CGCTGCAAGAGATGGCGTCAAGCAGGATGTATCCAGCAGCGCCGCGAGTGGTGAGT 397
DB 85 AlaSerSerArgProLeuProArgLeuArgThrLysSerProGluProThrSer 104
QY 398 -----CCACCGCGCAGCAATTTGACTTCCCA---ACAACTGCCGACCCCACTGTTT 448
DB 105 HisArgSerGlyAlaSerAsnProThrArgProLeuThrIleCysThrProGluGluThr 124
QY 449 CATTA-----ACCAACCCCGCCT-----GAGCGC 474
DB 125 HisProSerGluThrArgProSerProThrArgAlaThrAlaTrpSerLeuThrGluArg 144
QY 475 TCATTCATCGATTTTGACGGGAGCGGCGCAGTTGC-----510
DB 145 ArgHisHisArg-----CysHisArgProAspThrTrpIleSer 157
QY 511 -----CGAGCGCGCCAGGGGCGCCAGGAGCTGCAAAATCGTTGCCAGC 555
DB 158 LeuIleAspArgAlaIleLeuGlyPheAlaGlyAlaCysLeuAlaMetAlaSer 175

---SULT 12
ABG05131
ID ABG05131 standard; Protein: 264 AA.
XX AC ABG05131;
XX AC ABG05131;
DT 13-FEB-2002 (first entry)
XX Novel human diagnostic protein #5122.
DE Human: chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX Homo sapiens.
OS WO200175067-A2.
XX PD 11-OCT-2001.
XX 30-MAR-2001; 2001WO-US08631.
XX 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX

(HYSE-) HYSEQ INC.
XX Drmanac RT, Liu C, Tang YT;
XX WPI: 2001-639362/73.
DR N-PSDB; AAS69318.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
XX biodiversity
PS Claim 20; SEQ ID No 35490; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 264 AA:
Alignment Scores:
Pred. No.: 0.0265 Length: 264
Score: 105.00 Matches: 59
Percent Similarity: 35.75% Conservative: 15
Best Local Similarity: 28.50% Mismatches: 69
Query Match: 9.84% Indels: 64
DB: 22 Gaps: 12

US-09-920-953-2 (1-598) x ABG05131 (1-264)

QY 8 CTGACAGCGCGCCACACGACGACGAGCGGAGCGGAGTGGCGGCGCAAGAGC 67
DB 89 LeuGlnAlaGlySerThrAlaPro-----GlyAlaGlyThrProGly 102
QY 68 TGTGTTGATGACCTGGCGCGCAGAGGATGAAGCTGGCGTTTCACACCTTCTACGATA 127
DB 103 SerArgProThrTrpSerSerSerThrCysSerThrThrAlaProSerGlyArgAla 122
QY 128 AGGTGCTGGCTGACCGGAGCTGCTGCCCTTCTTCAGT-----166
DB 123 ArgCys-----AlaCysAlaSerSerSerCysAlaMetSerAlaAlaArg 138
QY 167 ---CCCTGGACATGCAAGAGC-----AGAGATGAAGCAGG-----TCAGATTCA 208
DB 139 ArgGlyTrpThrSerProAlaCysTrpArgArgThrSerArgAlaTrpThrThrSer 158
QY 209 TGAGCTTCGTGTTGGCGGAGCAGACCAATACAAGGGCGGAGCATGTACGACGACACG 268
DB 159 SerProAlaCysAlaSerSerAlaThrAlaSerValAlaAla-----172
QY 269 CCCATCTGCTCAAGGCGCCAGCGCTGGACCGCCACTTTTCACAGATCAAGCAGTACC 328
DB 173 -----SerThrAlaSerThrTrpProAlaAlaArgThrThr-----GlyGlyThr 187
QY 329 TTGGAGAGACGCTGCAAGAGATGGCGCTCAAGCAGGATGT-----GATCCAGCA-----377

CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 1307 AA;

Alignment Scores:

Pred. No.: 0.0697 Length: 1307
 Score: 103.50 Matches: 48
 Percent Similarity: 38.95% Conservative: 26
 Best Local Similarity: 25.26% Mismatches: 69
 Query Match: 9.70% Indels: 47
 DB: 22 Gaps: 9

US-09-920-953-2 (1-598) x ABB65464 (1-1307)

OY 56 GGCCGACAGAGCTTTGATGACCTGGCGCGCAGAGCATGAAGCTGGCGTTGACA 115
 Db 451 GlyMetSerThrTyValPrometSerMetAspSerSerGluSerLeuGlyLeuAsp 470
 OY 116 CTTTCTACGATAAG---TGCTGGCTGACCGCGGAGCTGCTCCCTCTTTCGAGTCCCTGG 172
 b 471 ProSerThrGluGluValIlePheMetThrAlaAsnArgThrProThrLeuGluProSer 490
 OY 173 ACATGCAAGACGACAGATGAAGCAGGTCAAGTTCATGAGCTTCGTGTTGGCGGAGCAG 232
 Db 491 ValThrSerSerLeuGlyThrSerSerSerAsnProAlaSerAlaIle---GlnPro 509
 OY 233 ACCAATACAGGGCGGAGGATGTACGACGACACGCCCATCTGCTCAAGGGCCACGCC 292
 Db 510 ThrSerSerLysGluHisThrLysThr----- 519
 OY 293 TGGACACCGCCACTTTGACAGATCAAGCAGTACCTTGGAGACACGCTGCAACAGATGG 352
 Db 520 ---ThrThrThrThrThrThrThrSerSerSerThr----- 530
 OY 353 GCGTCAAGCAGGATGTGATCCAGCAGCGCCCGGAGTGGTGAGTCCACCGCGACGAAT 412
 Db 531 -----SerThrThrThrThrThrProProSerThrGlu 543
 OY 413 TTGACTTCCCACTGCGCAC-----CCAACTGATTTTCAT 451
 Db 544 SerValSerProAsnSerSerAsnAlaAsnThrAsnAlaThrProProAlaProTyrHis 563
 OY 452 TAACCCAAACCCCGCTGAGCGCTATTCCATCGATTTTGAGCGGGGAGCG----- 502
 Db 564 ---ProArgProGlyIleValLeuAspAspProGluPheLysProGlyGlyArgProArg 582
 OY 503 CCAGTTGCCGAGCGCGCCAGGGGGCCAGGAGCGCTCAATCGT-----TTGCCA 553
 Db 583 ProProValGlnArg-----ProProAlaGlnGlnThrLeuProLeuPro 597
 JY 554 GCCCTTGTGATTGAGAGCCATCAGCCA 583
 Db 598 AlaValGlnProThrArgGlnHisLeuPro 607

Search completed: April 16, 2003, 12:58:50
 Job time : 43.5 secs

Db 83 ValGlnArgSerLysGlnPheAlaPheLeuAlaTyrAlaLeuAlaGlyAlaSerGluTrp 102

Qy 241 AAGGCCCAAGCATCTAGCAGCAGCACGCCCATCTGGTCAAGGCCACGGCCCTGGACCAC 300

Db 103 LysGlyLysAspMetArgThrAlaHisLysAspLeuValPro-----HisLeuSerAsp 120

Qy 301 GCCACTTTGACAGATCAAGCAGTACCTTGGAGAGAGCGTCAAGAGATGGCGCTCAAG 360

Db 121 ValHisPheGlnAlaValAlaArgHisLeuSerAspThrLeuThrGluLeuGlyValPro 140

Qy 361 CAGGATGTGATCCACCAGCCGCCCGAGTGGTGAGTCCACCCCGGACGAA 411

Db 141 ProGluAspIleThrAspAlaMetAlaValAlaValAlaSerThrArgThrGlu 157

RESULT 3

A36270

hemoglobin - Tetrahymena pyriformis

C:Species: Tetrahymena pyriformis

C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 16-Jun-2000

C:Accession: S32555; A36270

R:Takagi, T.; Iwaasa, H.; Yuasa, H.; Shikama, K.; Takemasa, T.; Watanabe, Y. Biochim. Biophys. Acta 1173 75-78, 1993

A:Title: Primary structure of Tetrahymena hemoglobins.

A:Reference number: S32555; MUID:93250050; PMID:8485156

A:Accession: S32555

A:Molecule type: mRNA

A:Residues: 1-121 <TAK>

A:Cross-references: EMBL:D13920; NID:g217409; PIDN:BA003015.1; PID:g418106

J. Biol. Chem. 265, 8603-8609, 1990

A:Title: Protozoan hemoglobin from Tetrahymena pyriformis. Isolation, characterisation, and primary structure.

A:Reference number: A36270; MUID:90256780; PMID:2111321

A:Accession: A36270

A:Molecule type: protein

A:Residues: 1-121 <IWA>

C:Genetics:

A:Genetic code: SGC5

C:Superfamily: Tetrahymena globin

C:Keywords: Chromoprotein; heme; iron; metalloprotein; monomer

F:1-121/Product: hemoglobin I #status experimental <MAT1>

F:3-121/Product: hemoglobin II #status experimental <MAT2>

F:73/Binding site: heme iron (His) (proximal axial ligand) #status predicted

Alignment Scores:

Pred. No.:	2,6e-14	Length:	121
Score:	242.00	Matches:	50
Percent Similarity:	60.87%	Conservative:	20
Best Local Similarity:	43.48%	Mismatches:	41
Query Watch:	22.68%	Indels:	4
DB:	1	Gaps:	1

US-09-920-953-2 (1-598) x A36270 (1-121)

Qy 67 CTGTTGTGATGACCTGGCGGCGCAGAGGCATAGAGCTGGGGTTGACACCTTCTACGAT 126

Db 7 IleTyrGluLysLeuGlyGlyGluAsnAlaMetLysAlaValProLeuPheTyrLys 26

Qy 127 AAGTGTGCTGGCTGACCGGACCTGCTGCCCTTCTTCGAGTCCCTGGACATCAAGAGCAG 186

Db 27 LysValLeuAlaAspGluArgValLysHisPheLysAsnThrAspMetAspHisGln 46

Qy 187 AAGATGAAGCAGGTCACGTTTCATGAGCTTCGTTGGCGGAGCAGACCAATACAGGCG 246

Db 47 ThrLysGlnGlnThrAspPheLeuThrMetLeuGlyGlyProAsnHisTyrLysGly 66

Qy 247 CGAAGCATGTACGACGACGACGCCCATCTGCTCAAGGGCCACGGCTCGACACCGCCAC 306

Db 67 LysAsnMetThrGluAlaHis-----LysGlyMetAsnLeuGlnAsnLeuHis 82

Qy 307 TTTGACAAAGATCAAGCATCTCTGGAGAGACGCTGCAAGAGATGGGCGTCAAGCAGGAT 366

Db 83 PheAspAlaIleLeuGluAsnLeuAlaAlaThrLeuLysGluLeuGlyValThrAspAla 102

Best Local Similarity: 34.40% Mismatches: 52
 Query Match: 17.62% Indels: 7
 DB: 2 Gaps: 2

US-09-920-953-2 (1-598) x C70761 (1-136)

QY 67 CTGTTTGTACCTGGCGGCGGCAAGGCATGAAGCTGGCGGTGACACCTTCTACGAT 126
 DB 15 IietyrAspLysIleGlyHisGluAlaIleGluValValGluAspPheTyrVal 34
 QY 127 AAGTGTCTGGTACCGGAGTGTCTTCTTCGAGTCCCTGGACATCAAGAGCAG 186
 DB 35 ArgValLeuAlaAspGlnLeuSerAlaPhePheSerGlyThrAsnMetSerArgLeu 54
 QY 187 AAGTGAAGCAGGTCAAGTTCATGCTTCTGTTTGGCGGAGCAGCAATCAAGGCG 246
 DB 55 LysGlyLysGlnValGluPhePheAlaAlaLeuGlyGlyProGluProTyrThrGly 74
 QY 247 CGAAGCATGTACGAGCGACACGCCCATCTGCTCAAGGCCCGCTGGACCGCCAC 306
 DB 75 AlaPrometLysGlnValHis-----GlnGlyArgGlyIleThrMetHis 90
 QY 307 TTTGACAGATCAAGCAGTACCTTGGAGAGACGCTGCAAGAGATGGCGTCAAGCAGGAT 366
 DB 91 PheSerLeuValAlaGlyHisLeuAlaAspAlaLeuThrAlaAlaGlyValProSerGlu 110
 QY 367 GTGATCCAGCAGCGCGGAGTGTGTGAGTCCACCGCGAGCAATTTGACTTNC----- 421
 DB 111 ThrIleThrGluIleLeuGlyValIle-AlaProLeuAlaValAspValThrSerGlyG1 130
 QY 422 -CCAACAACCTGCG 433

RESULT 7
 S60030

hemoglobin - Paramesidium jenningsi

C:Species: Paramesidium jenningsi

C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 16-Jun-2000

C:Accession: S60030

R:Yamauchi, K.; Tada, H.; Usuki, I.

A:Title: Structure and evolution of Paramesidium hemoglobin genes.

A:Reference number: S60030; MUID:96038820; PMID:7578257

A:Accession: S60030

A:Molecule type: DNA

A:Residues: 1-117 <YAM>

Cross-references: EMBL:D49689; NID:g1071656; PIDN:BAA08540.1; PID:g1384087

Note: the authors did not translate the codon for residue 1

Genetics:

A:Gene: Hb

A:Genetic code: SGC5

A:Introns: 63/73

C:Superfamily: Tetrahymena globin

C:Keywords: chromoprotein; heme; iron; metalloprotein

F:69/Binding site: heme iron (His) (proximal axial ligand) #status predicted

Alignment Scores:

Pred. No.: 1.12e-05 Length: 117

Score: 149.00 Matches: 31

Percent Similarity: 46.96% Conservative: 23

Best Local Similarity: 26.96% Mismatches: 57

Query Match: 13.96% Indels: 4

DB: 1 Gaps: 1

US-09-920-953-2 (1-598) x S60030 (1-117)

QY 67 CTGTTTGTACCTGGCGGCGGCAAGGCATGAAGCTGGCGGTGACACCTTCTACGAT 126

DB 3 LeuPheGluGlnLeuGlyGlyGluAlaAlaValThrAlaValThrGlnPheTyrAla 22

QY 127 AAGTGTCTGCTACCGGAGTGTCTTCTTCGAGTCCCTGGACATGCAAGAGCAG 186

DB 23 AsnIleGlnAlaAspAlaThrValAlaAsnPheAsnGlyIleAsnMetAlaAspGln 42

QY 187 AAGTGAAGCAGGTCAAGTTCATGCTTCTGTTTGGCGGAGCAGCAATCAAGGCG 246

DB 43 ThrAsnLysThrAlaSerPheLeuCysAlaAlaLeuGlyGlyProLysAlaTyrGlyGly 62

QY 247 CGAAGCATGTACGAGCGACACGCCCATCTGCTCAAGGCCCGCTGGACCGCCAC 306

DB 63 ArgAsnLeuLysGluValHisAlaAsnMet-----GlyValThrAsnAlaGln 78

QY 307 TTTGACAAGATCAAGCAGTACCTTGGAGAGACGCTGCAAGAGATGGCGTCAAGCAGGAT 366

DB 79 PheThrValIleGlyHisLeuArgSerAlaLeuThrSerAlaGlyValAlaAlaAsp 98

DB 23 AsnIleGlnAlaAspAlaThrValAlaAsnPheAsnGlyIleAsnMetAlaAspGln 42

QY 187 AAGTGAAGCAGGTCAAGTTCATGCTTCTGTTTGGCGGAGCAGCAATCAAGGCG 246

DB 43 ThrAsnLysThrAlaSerPheLeuCysAlaAlaLeuGlyGlyProLysAlaTyrGlyGly 62

QY 247 CGAAGCATGTACGAGCGACACGCCCATCTGCTCAAGGCCCGCTGGACCGCCAC 306

DB 63 ArgAsnLeuLysGluValHisAlaAsnMet-----GlyValThrAsnAlaGln 78

QY 307 TTTGACAAGATCAAGCAGTACCTTGGAGAGACGCTGCAAGAGATGGCGTCAAGCAGGAT 366

DB 79 PheThrValIleGlyHisLeuArgSerAlaLeuThrSerAlaGlyValAlaAlaAsp 98

QY 367 GTGATCCAGCAGCGCGCGGAGTGTGCGAGTCCACCGCGCGAGAA 411

DB 99 LeuValGluGlnThrValAlaValAlaGluThrValArgGlyAsp 113

RESULT 8

S60032

hemoglobin - Paramesidium triaurelia

C:Species: Paramesidium triaurelia

C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 16-Jun-2000

C:Accession: S60032

R:Yamauchi, K.; Tada, H.; Usuki, I.

A:Title: Structure and evolution of Paramesidium hemoglobin genes.

A:Reference number: S60030; MUID:96038820; PMID:7578257

A:Accession: S60032

A:Molecule type: DNA

A:Residues: 1-117 <YAM>

A:Cross-references: EMBL:D49688; NID:g1071655; PIDN:BAA08539.1; PID:g1384086

A:Note: the authors did not translate the codon for residue 1

Genetics:

A:Gene: Hb

A:Genetic code: SGC5

A:Introns: 63/73

C:Superfamily: Tetrahymena globin

C:Keywords: chromoprotein; heme; iron; metalloprotein

F:69/Binding site: heme iron (His) (proximal axial ligand) #status predicted

Alignment Scores:

Pred. No.: 1.12e-05 Length: 117

Score: 149.00 Matches: 31

Percent Similarity: 46.96% Conservative: 23

Best Local Similarity: 26.96% Mismatches: 57

Query Match: 13.96% Indels: 4

DB: 1 Gaps: 1

US-09-920-953-2 (1-598) x S60032 (1-117)

QY 67 CTGTTTGTACCTGGCGGCGGCAAGGCATGAAGCTGGCGGTGACACCTTCTACGAT 126

DB 3 LeuPheGluGlnLeuGlyGlyGluAlaAlaValThrAlaValThrGlnPheTyrAla 22

QY 127 AAGTGTCTGCTACCGGAGTGTCTTCTTCGAGTCCCTGGACATGCAAGAGCAG 186

DB 23 AsnIleGlnAlaAspAlaThrValAlaAsnPheAsnGlyIleAsnMetAlaAspGln 42

QY 187 AAGTGAAGCAGGTCAAGTTCATGCTTCTGTTTGGCGGAGCAGCAATCAAGGCG 246

DB 43 ThrAsnLysThrAlaSerPheLeuCysAlaAlaLeuGlyGlyProLysAlaTyrGlyGly 62

QY 247 CGAAGCATGTACGAGCGACACGCCCATCTGCTCAAGGCCCGCTGGACCGCCAC 306

DB 63 ArgAsnLeuLysGluValHisAlaAsnMet-----GlyValThrAsnAlaGln 78

QY 307 TTTGACAAGATCAAGCAGTACCTTGGAGAGACGCTGCAAGAGATGGCGTCAAGCAGGAT 366

DB 79 PheThrValIleGlyHisLeuArgSerAlaLeuThrSerAlaGlyValAlaAlaAsp 98

QY 367 GTGATCCAGCAGCGCGCGGAGTGTGCGAGTCCACCGCGCGAGAA 411

DB 99 LeuValGluGlnThrValAlaValAlaGluThrValArgGlyAsp 113

[illegible][illegible]


```

Db      168 yArgProGlnArgProValProArgProPheProGlyLeuGlnSerProGlyCysProPr 188
          |||||
Qy      300 -----CCGCCACTTTTGACAAGATCAAGACGTACTCTGGAG 334
          ||||||| ||| :||| ||| ||
Db      188 oGluGlyThrLeuGlyValProSerProProLeu---GlnAlaArgAlaSerProSerAr 207
          ||| ||| ||| ||| ||| |||
Qy      335 AGACGCTGCAAGATGGCGCTCAAGCAGGATGTCATCCAGCACGCCCGCGGACGTGGTGG 394
          ||| ||| ||| ||| ||| |||
Db      207 gArgGlyAlaSerLeuGlyProGlnVal-----GlnProHisArgAspProSerGlyPr 225
          ||| ||| ||| ||| ||| |||
Qy      395 -AGTCCACCCCGCAGGAATTGACTTNCCCACAACTCGGCACCACCACTGATTTTCATTTA 453
          ||||| ||| |||: ||| |||
Db      225 oAspProPro-----ThrGlyProSerLeuCysProproAlaProLeu----- 239
          ||||| ||| ||| ||| ||| |||
Qy      454 ACCCAACCCCCAGCCTGAGCGCTCATTTCCATCGATTGTCAGCGGGAGCGCCAGTTCGCCGA 513
          ||||||| ||| :|||: ||| |||
Db      240 ---GlnProSerLeuHisProArg-ProGlnLeuLeuAlaSerProGlyPro--ProG 257
          ||| |||||| | ||| |||||
Qy      514 GCGGCGCCAGGGGCGCCAGGACCTGCGAATCGTTTGGCAGCCCTTG 560
          ||| |||||| | ||| |||||
Db      257 lyGlnProGluGlyProArgGlnProGlyArgValAlaPheProLeu 272
          ||| ||||| ||| ||| |||||

RESULT 14
F84006
hypothetical protein BH2854 [imported] - Bacillus halodurans (strain C-125)
C:Species: Bacillus halodurans
C>Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
C:Accession: F84006
R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, I.
Nucleic Acids Res. 28, 4317-4331, 2000
A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans
A:Reference number: A83650; MUID:20512582; PMID:11058132
A:Accession: F84006
A:Molecule type: DNA
A>Status: preliminary
A:Cross-references: GB:APO01516; GB:BA000004; NID:g10175192; PIDN:BAB06573.1; GSI>
A:Experimental source: strain C-125
C:Genetics:
A:Gene: BH2854

Alignment Scores:
Pred. NO.: 0.171 Length: 130
Score: 104.00 Matches: 23
Percent Similarity: 50.70% Conservative: 13
Best Local Similarity: 32.39% Mismatches: 29
Query Match: 9.75% Indels: 6
DB: 2 Gaps: 2

US-09-920-953-2 (1-598) x F84006 (1-130)
Qy      61 AAGAGCTGTTTGATGACTCGTGGCGCGCAGAAGCATGAAGCTGCGGTTGACCTTC 120
          ::::: ||||| ||| ||| ||| ||| ||| |||
Db      6 ArgThrValTyGluArgIleGlyAspLysThrValserAlaLeuValaspAlaPhe 25
          ||| ||| ||| ||| ||| ||| ||| |||
Qy      121 TACGATAAGTGCTGGCTGACCGGAGCTGCTGCCCTTCTTCGAGTCCCTGGACATGCAA 180
          ||| ||| ||| ||| ||| ||| ||| |||
Db      26 TyrThrAsnValalaglyAspProIleLeuAlaProilePheProasn---AspLeuThr 44
          ||| ||| ||| ||| ||| ||| ||| |||
Qy      181 GAGCAGACATCAAGCAGCTCAAGTYTCATGACCTTCGTGTCGCGGAGCAGCACCAATAC 240
          ||| ||| ||| ||| ||| ||| ||| |||
Db      45 GluThrLysArgLysGlnGlnGlnPheLeuThrGlnPheLeuGlyGlyProSerLeuTy 64
          ||| ||| ||| ||| ||| ||| ||| |||
Qy      241 AAGGCGCGAAGCATGTACGACGCACACGCCCAT 273
          ||| ||| ||| ||| |||
Db      65 Thr-----GluGluHisGlyHis 70
          ||| ||| ||| ||| |||

RESULT 15
T35474
50kD proline rich protein - Streptomyces coelicolor
C:Species: Streptomyces coelicolor
C>Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 05-Nov-1999

```


Search completed: April 16, 2003, 13:01:58
Job time : 29 secs


```
OY 361 CAGGATGTGATCAGCAGCGCGGAGTGGTGGAGTCCACCGCAGCAA 411
Db 141 ProGluAspIleThrAspAlaMetAlaValAlaSerThrArgThrGlu 157

RESULT 3
GLB_TETPY
ID GLB_TETPY STANDARD; PRT; 121 AA.
AC P17724;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE Myoglobin (Hemoglobin).
OS Tetrahymena pyriformis.
OC Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Hymenostomatida;
OC Tetrahymenina; Tetrahymena.
OX NCBI_TaxID=5908;
RN [1]
RN MEDLINE=90256780; PubMed=2111321;
RX Iwasa H., Takagi T., Shikama K.;
YA Takagi T., Iwasa H., Yuasa H., Shikama K., Takemasa T.,
YA Watanabe Y.;
RT "Primary structure of Tetrahymena hemoglobins.";
RL Biochim. Biophys. Acta 1173:75-78(1993).
RN [2]
RN MEDLINE=90256780; PubMed=2111321;
RX Iwasa H., Takagi T., Shikama K.;
RT "Protozoan hemoglobin from Tetrahymena pyriformis. Isolation,
characterization, and amino acid sequence.";
RL J. Biol. Chem. 265:8603-8609(1990).
CC -1- SUBUNIT: MONOMER.
CC -1- SIMILARITY: BELONGS TO THE PROTOZOAN/CYANOBACTERIAL GLOBIN
FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
DR EMBL: D13920; BAA03015.1;
DR PIR: A36270; A36270.
DR PIR: S32555; S32555.
DR InterPro: IPR001486; Bac_globin.
DR Pfam: PF01152; Bac_globin; 1.
DR ProDom: PD004840; Bac_globin; 1.
R PROSITE: PS01213; GLOBIN_FAM_2; 1.
AW Heme; Oxygen transport; Transprot; Acetylation.
FT MOD_RES 1 1 ACETYLATION (PROBABLE).
FT METAL 73 73 IRON (HEME PROXIMAL LIGAND) (POTENTIAL).
SQ SEQUENCE 121 AA; 13684 MW; 6D02948060CA2A5E CRC64;

Alignment Scores:
Pred. No.: 3.43e-13 Length: 121
Score: 242.00 Matches: 50
Percent Similarity: 60.87% Conservative: 20
Best Local Similarity: 43.48% Mismatches: 41
Query Match: 22.68% Indels: 4
DB: 1 Gaps: 1

US-09-920-953-2 (1-598) x GLB_TETPY (1-121)

OY 67 CTGTTTGTATGACCTGGCGGCAGAGGATGAAGTGGCGTTGACACCTTCTACGAT 126
Db 7 IleTyrGluLysLeuGlyGlyGluAsnAlaMetLysAlaAlaValProLeuPheTyrLys 26

OY 127 AAGGTGCTGGCTGCCCGGACCTGCTGCCCTTCTTCGAGTCCCTGGACATCAAGACAG 186
Db 27 LysValLeuAlaAspGluArgValLysHisPhePheLysAsnThrAspMetAspHisGln 46
```

```
OY 187 AAGATGAAGCAGGTCAAGTTCATGAGCTTGTGGCGGAGCAGACCAATACAAGGC 246
Db 47 ThrLysGlnGlnThrAspPheLeuThrMetLeuLeuGlyGlyProAsnHisTyrLysGly 66

OY 247 CGAAGCATGTACGACGACGCCCATCTGTGTGAAGGCGCCAGCGCTGGACCCGCCAC 306
Db 67 LysAsnMetThrGluAlaHis-----LysGlyMetAsnLeuGlnAsnLeuHis 82

OY 307 TTTTCACCAAGATCAAGCAGTACCTTGGAGAGAGCGTGCACAGAGATGGCGGTCAAGCAGAT 366
Db 83 PheAspAlaIleIleGluAsnLeuAlaAlaThrLeuLysGluLeuGlyValThrAspAla 102

OY 367 GTGATCCAGCAGCGCGGAGTGGTGGAGTCCACCCCGCAGCAA 411
Db 103 ValIleAsnGluAlaAlaLysValIleGluHisThrArgLysAsp 117

RESULT 4
GLB2_CHLEU
ID GLB2_CHLEU STANDARD; PRT; 171 AA.
AC P52334;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Globin Li410.
GN Li410.
OS Chlamydomonas eugametos.
OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
OC Chlamydomonadaceae; Chlamydomonas.
OX NCBI_TaxID=3053;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=UTEX 9;
RX MEDLINE=94232186; PubMed=8177215;
RA Couture M., Chamberland H., St Pierre B., Lafontaine J., Guertin M.;
RT "Nuclear genes encoding chloroplast hemoglobins in the unicellular
green alga Chlamydomonas eugametos.";
RL Mol. Gen. Genet. 243:185-197(1994).
CC -1- SUBCELLULAR LOCATION: CHLOROPLAST, PARTICULARLY IN THE PYRENOID
CC AND THE THYLAKOID REGION.
CC -1- INDUCTION: BY LIGHT.
CC -1- SIMILARITY: BELONGS TO THE PROTOZOAN/CYANOBACTERIAL GLOBIN
FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
DR EMBL: X72915; CAA51420.1;
DR InterPro: IPR001486; Bac_globin.
DR Pfam: PF01152; Bac_globin; 1.
DR ProDom: PD004840; Bac_globin; 1.
DR PROSITE: PS01213; GLOBIN_FAM_2; 1.
KW Heme; Oxygen transport; Transprot; Chloroplast.
FT METAL 63 63 IRON (HEME DISTAL LIGAND) (BY
SIMILARITY).
FT METAL 111 111 IRON (HEME PROXIMAL LIGAND) (POTENTIAL).
SQ SEQUENCE 171 AA; 18598 MW; 236E759D0C2D42FA CRC64;

Alignment Scores:
Pred. No.: 3.9e-13 Length: 171
Score: 241.50 Matches: 56
Percent Similarity: 57.58% Conservative: 20
Best Local Similarity: 42.42% Mismatches: 53
Query Match: 22.63% Indels: 3
DB: 1 Gaps: 2

US-09-920-953-2 (1-598) x GLB2_CHLEU (1-171)

OY 16 GCGGCCACCCAGCACCAGACGCGGGAAGCGGATCGGGCGCAGACGCTGTTGAT 75
```

```

Best Local Similarity: 40.87%      Mismatches:      41
Query Match:      21.56%      Indels:      4
DB:      1      Gaps:      1

US-09-920-953-2 (1-598) x GLB_TETH (1-121)

QY 67 CTGTTTGATGACCTGGCGGCGCAGAAAGCATGACCTGGCGGTTTGACACCTTCACGAT 126
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 7 ValPheGluLysLeuGlyGlnAlaAlaMetHisAlaAlaValProLeuPheTyrLys 26
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
QY 127 AAGGTGCTGCTCACCGGAGCTGCTCTTCTTCAGTGCCTCGTGGACATCAAGAGCAG 186
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 27 LysValLeuAlaAspArgValLysHisTyrPheLysAsnThrAsnMetGluHisGln 46
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
QY 187 AAGATGAAGCAGCTCAAGTTCATGAGCTTCGCTGTTGGCGGAGCAGCAACATACAAGGC 246
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 47 AlaLysGlnGlnGluAspPheLeuThrMetLeuGluGlyGlyProAsnHisTyrLysGly 66
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
QY 247 CGAAGCATGTACGACGACACCGCCCATCTGCTCAGGGCCGACGGCTGACACCCGCCAC 306
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 67 LysAsnMetAlaGluAlaHis-----LysGlyMetAsnLeuGlnAsnSerHis 82
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
QY 307 TTTGACAAGATCAAGCAGTACCTTGGGAGAGAGCTGCAAGAGATGGCGCTCAAGCAGGAT 366
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 83 PheAspAlaIleIleGluAsnLeuAlaAlaThrLeuLysGluLeuGlyValSerAspGln 102
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
QY 367 GTGATCCAGCAGCCCGCGGAGTGTGGAGTGCACCGCGCAGGAA 411
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 103 IleIleGlyGluAlaAlaLysValIleGluHisThrArgLysAsp 117
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::

RESULT 6
      GLBN_MYCTU
ID      GLBN_MYCTU      STANDARD;      PRT;      136 AA.
AC      Q10784;
DT      01-OCT-1996 (Rel. 34, Created)
DT      01-OCT-1996 (Rel. 34, Last sequence update)
DT      13-JUN-2002 (Rel. 41, Last annotation update)
DE      Hemoglobin-like protein HbN (Flavohemoglobin).
GN      GLBN OR RV1542 OR MT1594 OR MTCY48.23.
OS      Mycobacterium tuberculosis,
OS      Mycobacterium bovis, and
OS      Mycobacterium smegmatis.
OC      Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC      Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
ON      NCBI_TaxID=1773, 1765, 1772;
RX      [1]
RC      SPECIES FROM N.A.
RC      SPECIES=M.tuberculosis; STRAIN=H37Rv;
RC      MEDLINE=98295987; PubMed=9634230;
RC      Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA      Gordon S.V., Eigmler K., Gas S., Barry C.E. III, Tekala F.,
RA      Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA      Davies K., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA      Horsby T., Jorgels K., Krohn A., McLean J., Moule S., Murphy L.,
RA      Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA      Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA      Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT      "Deciphering the biology of Mycobacterium tuberculosis from the
RT      complete genome sequence.";
RL      Nature 393:537-544(1998).
RL      [2]
RC      SPECIES FROM N.A.
RC      SPECIES=M.tuberculosis; STRAIN=CDC 1551 / Oshkosh;
RA      Felschmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA      Petersson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA      Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA      Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA      Bishai W.;
RT      "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT      laboratory strains";
RL      Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
RL      [3]
RC      SPECIES FROM N.A., AND CHARACTERIZATION.
RN      RP

```

RC SPECIES=M.bovis; STRAIN=BCG;
RX MEDLINE=99432214; PubMed=10500158;
RA Couture M., Reh S.R., Wittenberg B.A., Wittenberg J.B., Ouellet Y.,
R Rousseau D.L., Guertin M.;
RT "A cooperative oxygen-binding hemoglobin from Mycobacterium
tuberculosis";
RL Proc. Natl. Acad. Sci. U.S.A. 96:11223-11228(1999).
RN [4]
RP SEQUENCE FROM N.A.
RC SPECIES=M.smegmatis; STRAIN=LR222;
RA Kumar A., Rawat V.P.S., Das R.H.;
RT "Cloning, expression and characterization of globin like genes of
Mycobacterium in insect cells(sf9).";
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
RN [5]
RP CHARACTERIZATION.
RX MEDLINE=20102674; PubMed=10636862;
RA Yeh S.R., Couture M., Ouellet Y., Rousseau D.L.;
RT "A cooperative oxygen binding hemoglobin from Mycobacterium
tuberculosis. Stabilization of heme ligands by a distal tyrosine
residue.";
J Biol. Chem. 275:1679-1684(2000).
CC -I- FUNCTION: Binds oxygen cooperatively with very high affinity
(P50) = 0.013 mmHg at 20 degrees Celsius) because of a fast
combination (25 microm(-1).s(-1)) and a slow dissociation (0.2 s(-
1)) rate.
CC -I- SUBUNIT: Homodimer.
CC -I- SIMILARITY: BELONGS TO THE PROTOZOAN/CYANOBACTERIAL GLOBIN
FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
or send an email to license@isb-sib.ch).
CC -----
CC EMBL: 274020; CAA98320.1; .
DR EMBL: AE007025; AAK45860.1; .
DR EMBL: AF130980; AAD28758.1; .
DR EMBL: AJ249386; CAB56291.1; .
DR TIGR: WU1594; .
DR TubercuList; RV1542c; .
DR InterPro: IPR001486; Bac_globin.
DR Pfam: PF01152; Bac_globin; 1.
DR ProDom: PD004840; Bac_globin; 1.
DR PROSITE: PS01213; GLOBIN_FAM_2; 1.
W Heme; Oxygen transport; Transport; Complete proteome.
I METAL 81 81 IRON (HEME PROXIMAL LIGAND) (POTENTIAL).
SQ SEQUENCE 136 AA; 1448 MW; B75D01A45BC0645B CRC64;

Alignment Scores:
Pred. No.: 1,48e-08 Length: 136
Score: 188.00 Matches: 43
Percent Similarity: 52.80% Conservative: 23
Best Local Similarity: 34.40% Mismatches: 52
Query Match: 7 Indels: 7
DB: 1 Gaps: 2

US-09-920-953-2 (1-598) x GLBN_MYCTU (1-136)

QY 67 CTGTTTGTACCTGGCGGCGGAGAGGATGAGTGGGGTTGACCTTCTACGAT 126
Db 15 IletyrAspLysIleGlyHisGluAlaIleGluValValGluAspPheTyrVal 34
QY 127 AAGTGTGCTGACCCCGGAGCTGCTCCCTTCTTCAGTCCCTGGACATCCAGACGAC 186
Db 35 ArgValLeuAlaAspAspGlnLeuSerAlaPhePheSerGlyThrAsnMetSerArgLeu 54
QY 187 AAGATGAGCAGGTCAAGTTTCATGAGCTTCGTGTTTGGCGGAGCAGACCAATACAAAGGC 246
Db 55 LysGlyLysGlnValGluPhePheAlaAlaLeuGlyGlyProGluProTyrThrGly 74

QY 247 CGAAGCATGTACGACGACACGCCCATCTGTGTCAGGCGGCGGCTGGACCCGCCAC 306
Db 75 AlaProMetLysGlnValHis-----GlnGlyArgGlyIleThrHethHis 90
QY 307 TTTGACAAGATCAACAGCTACCTTGGAGAGAGCTGCAAGAGATGGCGCTCAAGCAGAT 366
Db 91 PheSerLeuValAlaGlyHisLeuAlaAspAlaLeuThrAlaAlaGlyValProSerGlu 110
QY 367 GTGATCCAGCAGCGCGGAGTGGTGGAGTCCACCGGACGAAATTTGACTTNC----- 421
Db 111 ThrIleThrGluIleLeuGlyValIle-AlaProLeuAlaValAspValThrSerGlyGlu 130
QY 422 -CCAACAACCTGGC 433
Db 130 uSerThrThrAla 134

RESULT 7
GLBN_NOSSN STANDARD; PRT; 118 AA.
AC P52335;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Cyanogloblin.
GN GLBN
OS Nostoc sp. (strain MUN 8820).
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX NCBI_TaxID=55397;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97086627; PubMed=8932316;
RA Hill D.R., Belbin T.J., Thorsteinsson M.V., Bassam D., Brass S.,
Ernst A., Boger P., Paerl H., Mulligan M.E., Potts M.;
RT "GLN (cyanogloblin) is a peripheral membrane protein that is
restricted to certain Nostoc spp.";
RL J. Bacteriol. 178:6587-6598(1996).
CC -I- SUBUNIT: MONOMER.
CC -I- SUBCELLULAR LOCATION: PERIPHERAL MEMBRANE PROTEIN.
CC -I- SIMILARITY: BELONGS TO THE PROTOZOAN/CYANOBACTERIAL GLOBIN
FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
or send an email to license@isb-sib.ch).
CC -----
CC EMBL: L47979; AAB41122.1; .
DR InterPro: IPR001486; Bac_globin.
DR Pfam: PF01152; Bac_globin; 1.
DR ProDom: PD004840; Bac_globin; 1.
DR PROSITE: PS01213; GLOBIN_FAM_2; 1.
KW Heme; Oxygen transport; Transport.
FT METAL 70 70 IRON (HEME PROXIMAL LIGAND) (POTENTIAL).
SQ SEQUENCE 118 AA; 12794 MW; D5B3E461FE4C860C CRC64;

Alignment Scores:
Pred. No.: 1.37e-06 Length: 118
Score: 165.00 Matches: 39
Percent Similarity: 45.45% Conservative: 16
Best Local Similarity: 32.23% Mismatches: 52
Query Match: 15.46% Indels: 14
DB: 1 Gaps: 2

US-09-920-953-2 (1-598) x GLBN_NOSSN (1-118)

QY 67 CTGTTTGTACCTGGCGGCGGAGAGGATGAGTGGGGTTGACCTTCTACGAT 126
Db 4 LeuTyrAspAsnIleGlyGlyGlnProAlaIleGluValValAspGluLeuHisLys 23

```

OY 127 AAGTCTGCTGGTCCACCGGAGCTGCTGCTCCCTTCGAGTCCCTCGACATGCAAGAGCAG 186
Db 24 ArgilleaThraspserLeuLeuAlaProIlePheAlaGlyThrAspMetAlaLysGln 43
OY 187 AAGTGAAGCAGCTCAAGTTCATGAGTTCGTTGGTGGGAGCAGACCAATACAGGCG 246
Db 44 ArgAsnHisLeuValAlaPheLeuGlyGlnIlePheGluGlyProLysGlnTyrGlyGly 63
OY 247 CGAAGCATGACGACGACACGCCCATCTGTTCAAGGGCCACGGCTGGACCCGCCAC 306
Db 64 ArgProMetAspLysThrHisAla-----GlyLeuAsnLeuGlnGlnProHis 79
OY 307 TTTGACAAGATCAACAGTACCTTGGAGAGACCTTGCACAGACATGGGCTCAAGCAGGAT 366
Db 80 PheAspAlaIleAlaLysHisLeuGlyGluAlaMetAlaValArgLysValSer----- 97
OY 367 GTGATCCAGCAGCGCGGAGTGGTGGAGTCCACCGGACGGAATTTGACTTNCCTCAAC 426
Db 98 -----AlaGluAspThrLysAlaAlaLeuAspArgValThr 109
427 AAC 429
111
110 Asn 110

RESULT 8
GLBN_NOSCO
ID GLBN_NOSCO STANDARD; PRT; 118 AA.
AC Q00812;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Cyanoglobin.
GN GLBN.
OS Nostoc commune.
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX NCBI_TaxID=1178; [1]
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=UTEX 584;
RX MEDLINE=92302835; PubMed=1609281;
RA Potts M., Angeloni S.V., Ebel R.E., Bassam D.;
RT "Myoglobin in a cyanobacterium.";
RL Science 256:1690-1692(1992). [2]
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=UTEX 584;
RX MEDLINE=94341563; PubMed=8063099;
RA Angeloni S.V., Potts M.;
RT "Analysis of the sequences within and flanking the cyanoglobin-
encoding gene, glbn, of the cyanobacterium Nostoc commune UTEX 584.";
Gene 146:133-134(1994).
CC -1- SUBUNIT: MONOMER.
CC -1- SUBCELLULAR LOCATION: PERIPHERAL MEMBRANE PROTEIN.
CC -1- SIMILARITY: BELONGS TO THE PROTOZOAN/CYANOBACTERIAL GLOBIN
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation :
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M92437; AAA25512.1; -
CC EMBL; L23514; AAA21837.1; -
CC InterPro: IPR001486; Bac-globin.
CC Pfam: PF01152; Bac-globin; 1.
CC ProDom: PD004840; Bac-globin; 1.
CC PROSITE: PS01213; GLOBIN_FAM_2; 1.
CC Heme; Oxygen transport; Transport.
CC METAL 70 70 IRON (HEME PROXIMAL LIGAND) (POTENTIAL).
CC CONFLICT 34 34 V -> I (IN REF. 2).

```

```

FT CONFLICT 41 52 V -> A (IN REF. 2).
FT CONFLICT 52 90 A -> G (IN REF. 2).
FT CONFLICT 90 100 R -> A (IN REF. 2).
FT CONFLICT 100 120 N -> D (IN REF. 2).
SQ SEQUENCE 118 AA; 12906 MW; 8202E4F93A1015E1 CRC64;

Alignment Scores:
Pred. No.: 1.67e-06 Length: 118
Score: 164.00 Matches: 39
Percent Similarity: 47.11% Conservative: 18
Best Local Similarity: 32.23% Mismatches: 50
Query Match: 15.37% Indels: 14
DB: 1 Gaps: 2

US-09-920-953-2 (1-598) x GLBN_NOSCO (1-118)
OY 67 CTGTTTGATGACCTGGCGGCGCAGAGGCATGAAGCTGGCGGTTGACACCTTCTACGAT 126
Db 4 LeutyAspAsnIleGlyGlnProAlaIleGluGlnValValAspGluLeuHisLys 23
OY 127 AAGTGTGCTGCTGACCCGAGCTGCTGCCCTTCTCGAGTCCCTCGAGATGCAAGAGCAG 186
Db 24 ArgIleAlaThrAspSerLeuLeuAlaProValPheAlaGlyThrAspMetValLysGln 43
OY 187 AAGATGAAGCAGCTCAAGTTCATGAGTTCGTTGGTGGGAGCAGACCAATACAGGCGC 246
Db 44 ArgAsnHisLeuValAlaPheLeuAlaGlnIlePheGluGlyProLysGlnTyrGlyGly 63
OY 247 CGAAGCATGACGACGACACGCCCATCTGTTCAAGGGCCACGGCTGGACCCGCCAC 306
Db 64 ArgProMetAspLysThrHisAla-----GlyLeuAsnLeuGlnGlnProHis 79
OY 307 TTTGACAAGATCAACAGTACCTTGGAGAGACCTTGCACAGACATGGGCTCAAGCAGGAT 366
Db 80 PheAspAlaIleAlaLysHisLeuGlyGluArgMetAlaValArgLysValSer----- 97
OY 367 GTGATCCAGCAGCGCGGAGTGGTGGAGTCCACCGGACGGAATTTGACTTNCCTCAAC 426
Db 98 -----AlaGluAsnThrLysAlaAlaLeuAspArgValThr 109
427 AAC 429
111
110 Asn 110

RESULT 9
GLB_PARCA
ID GLB_PARCA STANDARD; PRT; 116 AA.
AC P15160;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Myoglobin (hemoglobin).
OS Paramacium caudatum.
OC Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Penicillida;
OC Paramacium.
OC NCBI_TaxID=5885;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92118012; PubMed=1731779;
RA Yamauchi K., Mukai M., Ochiai T., Usuki I.;
RT "Molecular cloning of the cDNA for the major hemoglobin component
from Paramacium caudatum.";
RL Biochem. Biophys. Res. Commun. 182:195-200(1992). [2]
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=93042011; PubMed=1420365;
RA Yamauchi K., Ochiai T., Usuki I.;
RT "The unique structure of the Paramacium caudatum hemoglobin gene: the
presence of one intron in the middle of the coding region.";
RL Biochim. Biophys. Acta 1171:81-87(1992). [3]
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=93246250; PubMed=8482540;

```

RA Yamauchi K., Tada H., Ochiai T., Usuki I.;
 RT "Structure of the Paramesium caudatum gene encoding the B-type of the
 RL major hemoglobin component.";
 RN Gene 126:243-246(1993).
 RP SEQUENCE.
 RX MEDLINE=89362481; PubMed=2769763;
 RA Iwasa H., Takagi T., Shikama K.;
 RT "Protozoan myoglobin from Paramesium caudatum. Its unusual amino acid
 sequence.";
 RL J. Mol. Biol. 208:355-358(1989).
 CC -|- SUBUNIT: MONOMER.
 CC -|- SIMILARITY: BELONGS TO THE PROTOZOAN/CYANOBACTERIAL GLOBIN
 CC FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: S49254; AAB24268.2;
 DR EMBL: M99047; AAB29447.2;
 DR EMBL: M57542; AAB29446.2;
 DR EMBL: D12916; BAA02300.1;
 DR PIR: S05230; S05230.
 DR PIR: S27185; S27185.
 DR PIR: JQ2009; JQ2009.
 DR InterPro: IPR001486; Bac_globin.
 DR Pfam: PF01152; Bac_globin; 1.
 DR ProDom: PD004840; Bac_globin; 1.
 DR PROSITE: PS01213; GLOBIN_FAM_2; 1.
 KW Heme: Oxygen transport; Transport; Acetylation.
 FT INIT MET 0
 FT MOD_RES 1 1 ACETYLATION.
 FT METAL 68 68 IRON (HEME PROXIMAL LIGAND) (POTENTIAL).
 FT CONFLICT 17 17 A -> D (IN REF.2).
 FT CONFLICT 17 17 A -> D (IN REF.2).
 SQ SEQUENCE 116 AA: 11907 MW: 4646AE315F44855 CRC64;
 Alignment Scores:
 Pred. No.: 4.78e-05 Length: 116
 Score: 147.00 Matches: 31
 Percent Similarity: 46.96% Conservative: 23
 Best Local Similarity: 26.96% Mismatches: 57
 Query Match: 13.78% Indels: 4
 DB: 1 Gaps: 1
 3-09-920-953-2 (1-598) x GLE_PARCA (1-116)
 QY 67 CTGTTGTGACCTGGCGGCGCAGAGCATGAAGCTGGCGGTGACACCTCTTACGAT 126
 Db 2 LeuPheGluGlnLeuGlyGlyGlnAlaAlaValGlnAlaValThrAlaGlnPheTyrAla 21
 QY 127 AAGTCTGCTGCTACCGGAGCTGCTGCTTCTTCTGAGTCCCTGGACATGCAAGAGCAG 186
 Db 22 AsnIleGlnAlaAspAlaThrValAlaThrPhePheAsnGlyIleAspMetProAsnGln 41
 QY 187 AAGATCAAGCAGCTCAAGTTTCATGAGCTTCTGTTGGCGGAGCAGACCAATACAAAGGC 246
 Db 42 ThrAsnLysThrAlaAlaPheLeuCysAlaAlaLeuGlyGlyProAsnAlaTTPThrGly 61
 QY 247 CGAAGCATGTCAGCAGCAGCAGCCCATCTGTCGAGGCCAGCGCTGGACACCGCCAC 306
 Db 62 ArgAsnLeuLysGluValHisAlaAsnMet-----GlyValSerAsnAlaGln 77
 QY 307 TTTTCAAGATCAAGCATGACCTTGGAGACGCTGCAAGAGATGGCGTCAAGCAGGAT 366
 Db 78 PheThrThrValIleGlyHisLeuArgSerAlaLeuThrGlyAlaGlyValAlaAla 97
 QY 367 GTGATCCAGCACCGCCCGGAGTGGTGAGTCCACCGCCGACGAA 411

Db 98 LeuValGluGlnThrValAlaValAlaGluThrValArgGlyAsp 112
 RESULT 10
 IE18_PVKKA STANDARD: PRT: 1446 AA.
 AC P33479;
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 01-FEB-1994 (Rel. 28, Last annotation update)
 DE Immediate-early protein IE180.
 GN IE.
 OS Pseudorabies virus (strain Kaplan) (PRV).
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
 OC Alphaherpesvirinae; Varicellovirus.
 OX NCBI_TaxID=33703;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91021039; PubMed=2171211;
 RA Vitek C., Kozmik Z., Paces V., Schirm S., Schwyz M.;
 RT "Pseudorabies virus immediate-early gene overlaps with an oppositely
 RT oriented open reading frame: characterization of their promoter and
 RT enhancer regions.";
 RL Virology 179:365-377(1990).
 CC -|- FUNCTION: THIS IE PROTEIN IS A MULTIFUNCTIONAL PROTEIN CAPABLE
 CC OF MIGRATING TO THE NUCLEUS, BINDING TO DNA, TRANS-ACTIVATING
 CC OTHER VIRAL GENES, AND AUTOREGULATING ITS OWN SYNTHESIS.
 CC -|- SUBCELLULAR LOCATION: NUCLEUS OF INFECTED CELLS.
 CC -|- PTM: A LONG STRETCH OF SERINE RESIDUES MAY BE A MAJOR SITE OF
 CC PHOSPHORYLATION.
 CC -|- SIMILARITY: BELONGS TO THE ICP4/IE140/IE180 FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: M34651; AAB47470.1;
 DR PIR: A45344; A45344.
 DR InterPro: IPR005205; Herpes_ICP4_C.
 DR InterPro: IPR005206; Herpes_ICP4_N.
 DR Pfam: PF03584; Herpes_ICP4_N; 1.
 DR Pfam: PF03585; Herpes_ICP4_C; 1.
 KW Early protein; Transcription regulation; Trans-acting factor;
 FT DNA-binding; Phosphorylation; Nuclear protein.
 FT DOMAIN 347 354 POLY-SER.
 FT DOMAIN 379 397 POLY-SER.
 SQ SEQUENCE 1446 AA: 148640 MW: 81F43A3DE3DDA068 CRC64;
 Alignment Scores:
 Pred. No.: 0.177 Length: 1446
 Score: 106.50 Matches: 67
 Percent Similarity: 31.07% Conservative: 20
 Best Local Similarity: 23.93% Mismatches: 51
 Query Match: 9.98% Indels: 142
 DB: 1 Gaps: 12
 US-09-920-953-2 (1-598) x IE18_PVKKA (1-1446)
 QY 23 CCAGCACACAGAGCGCGGAGCGGATGCGGCGCAAGAAGCTGTTTGTGACCTCG 82
 Db 879 ProlAPrGluArgArgLysLysLys-ArgArgAla-----ProG 892
 QY 83 CGCGCGCAGAGCATGAAGCTGGCGGTGACACCTTCTACGATAAGTCTCGCTGACC 142
 Db 892 yAlaArgArg-----Pr 896
 QY 143 CGGA-----GCTGCTGCCCTTCTTCGAGTCCCTCGACATGC 178
 Db 896 oGlyAspGlyGluGluAspGluGlyLeuSerGlyAlaAlaLeuArgGlyAspGlyHisG 916


```

OY 179 AAGACGACGAGTGAAGCAGGTCAAGTTTCATGACGCTTCCTGTTGGCGGACGACGACCAAT 238
DB 916 yHisArgAspGluGluAsp----- 923
OY 239 ACAAGGCGCGAGCATGTACGACGACACAGCCCATCTGTCGAAGGCGCCACGCGCTGGACC 298
DB 924 -ArgGlyProArgGlyArgSerLeuGlyLeuGlyProAlaProAspProAlaP 943
OY 299 ACCGCCACTTGCACAGATCAAGCAGTACCTGGAG----- 334
DB 943 o--AlaLeuLeuSerSerSerSerSerGluAspAspArgLeuArgArgProLeu 962
OY 335 -----AGACCC----- 340
DB 963 GlyProMetProGluHisProAlaProAspGlyGlyPheArgArgValProAlaGlyGlu 982
OY 341 -----TCCAGAGATGGCGTCAAGCAGGATGTGATCCAGCAGCGC----- 382
DB 983 ThrHisThrProArgProSerGluAlaAlaLeuAlaAlaTyrCysProProGluValAla 1002
DB 383 -----CCGAGTGGTGGAGTCCACCCGCGACGAAATTG 415
DB 1003 ArgAlaLeuValAspGlnGluValPheProGluLeuTrpArgProAlaLeuThrPhe--- 1021
OY 416 ACTNCCCAACAACGCGCCCACTGATTTTCAATTAACCCACCCAGCCCTGAGCGCT 475
DB 1022 -----AspProAlaAlaLeuAla 1027
OY 476 CATTCCATCGATTGTCAGCGGGAGCGCCAGTT----- 508
DB 1028 His---IleAlaAlaArgArgGlyAlaProLeuArgArgAlaAlaTrpMetArgGln 1046
OY 508 ----- 508
DB 1047 IleAlaAspProGluAspValArgValValLeuTyrAspProLeuProHisGluGlu 1066
OY 509 -----GCCAGCGCGCCAGGCGCCAGGAGCGCTCA-----AATCGT 547
DB 1067 LeuGlyAlaGluProAlaGluGlyAlaProArgProAlaTrpAspProArgArgGlyGly 1086
OY 548 TTCCAGCCCTTCTGCTGATGAAGAGCCATCAG-----CCATTTTGGCAC 592
DB 1087 LeuSerAlaLeuAlaAlaPheAlaHisArgLeuGlyCysThrProAspSerHis 1104
RESULT 11
ID HB9_HUMAN STANDARD; PRT: 401 AA.
AC P50219;
O1-OCT-1996 (Rel. 34, Created)
O1-OCT-1996 (Rel. 34, Last sequence update)
15-JUN-2002 (Rel. 41, Last annotation update)
DE Homeobox protein HB9.
GN HLXB9.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE=Placenta;
RC MEDLINE=94327547; PubMed=7914194;
RA Harrison K.A., Druey K.M., Deguchi Y., Tuscano J.M., Kehrl J.H.;
RT "A novel human homeobox gene distantly related to proboscipedia is
RL J. Biol. Chem. 269:19968-19975(1994).
CC -!- FUNCTION: PUTATIVE TRANSCRIPTION FACTOR.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- TISSUE SPECIFICITY: EXPRESSED IN LYMPHOID AND PANCREATIC TISSUES.
CC -!- SIMILARITY: CONTAINS 1 HOMEBOX DOMAIN.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
```

```

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement. (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U07664; AAB60647.1; -.
DR EXBL; U07663; AAB60647.1; JOINED.
DR HSP; P14653; I872.
DR TRANSFAC; T03420; -.
DR Genew; HGNC:4979; HLXB9.
DR MIM; 142994; -.
DR InterPro; IPR001356; Homeobox.
DR Pfam; PF00046; homeobox; 1.
DR PRINTS; PR00024; HOMEBOX.
DR ProDom; PD000010; Homeobox; 1.
DR SMART; SM00389; HOX; 1.
DR PROSITE; PS00027; HOMEBOX_1; 1.
DR PROSITE; PS00071; HOMEBOX_2; 1.
DR Homeobox; DNA-binding; Nuclear protein; Transcription regulation.
KW DOMAIN 39 48 POLY-GLY.
FT DOMAIN 97 111 POLY-GLY.
FT DOMAIN 120 135 POLY-GLY.
FT DOMAIN 169 177 POLY-ALA.
FT DNA_BIND 242 301 HOMEBOX.
FT DOMAIN 316 325 POLY-GLY.
SQ SEQUENCE 401 AA; 40932 MW; 0006AED71D594FE CRC64;

Alignment Scores:
Pred. No.: 2.51 Length: 401
Score: 92.50 Matches: 44
Percent Similarity: 34.97% Conservative: 6
Best Local Similarity: 30.77% Mismatches: 39
Query Match: 8.24% Indels: 54
DB: 1 Gaps: 9

US-09-920-953-2 (1-598) x HB9_HUMAN (1-401)
OY 436 GTCCGCGAGTGTGGGNAAGTCAAATTCGTGCGGGTGACT-----CCACCACTC 386
DB 9 IleGluProCysTrpArg-----TrpThrProHisGluProProLeu 22
OY 385 CGCGCGGTGCTGGATCACATCTCTGACGCCCATCTCTTGCAGCGCTCTCTCCAAGT 326
DB 23 AlaGluArgAlaLeuAlaLysValThr---SerProValProAlaSer-----Gly 39
OY 325 ACTGCTTGATCTGTCAAAGTGGCGGTGTCAGCGCGTGGCCCTTGACCAAGATGGCGCT 266
DB 40 Thr-----GlyGlyGlyGlyGlyGlyGly-----AlaSerGlyGly 52
OY 265 GTGCGTGTACATGCTTCGCGCCCTTCTATTGCTCTCTCCGCCAAACACCAAGCTCATGA 206
DB 53 ThrSerGlySerCysSer----- 58
OY 205 ACTTGACCTGCTTCATCTTCTGCTCTTGCATGTCAGGACTCGAAGAGGGCAGCAGCT 146
DB 59 -----ProAlaSerSerGluProProAlaAlaProAlaAspArgLeuArgAlaGluSer 76
OY 145 CCGGGTCACCCAGCACCTTATCGTAGAAGGTGTCAACCGCCAGCTTCATGCTTCGCGC 86
DB 77 ProSerProProArg-----LeuLeu--- 83
OY 85 CGCCGAGGTATCAAAACAGCTTCTTGGCGCCGATCCCGCTCCGCGCTCTCTGTGGTGC 26
DB 84 -----AlaAlaHisCysAlaLeuLeuProLysProGlyPheLeu-GlyAl 98
OY 25 TGTGGC 19
DB 98 aclyGly 100
RESULT 12
ID UPN_MOUSE
AC Q90ZL6; STANDARD; PRT: 566 AA.
```


Db 1076 Met-----AspAsnLeuLysAsnSerArgLeuAlaThrAlaGluProVal 1090
QY 378 CGCGCGGGAGTGGTGGAGTCCACCGCGGACGAATTGACTTNCACCAACTCGCGCAC 437
Db 1091 SerProHisGluAsnLeuSerHisAlaGlyLeuProGlnSerProAlaLysMetGlySer 1110
QY 438 CAACACTATTTCATTAAACCAACCCGAGCTGAGCGCTCATTCATCGATTTGACGGCG 497
Db 1111 SerThrAspProAlaGlyProThrProAlaThrAlaAlaAsnProGlnAsnSerThrAla 1130
QY 498 GAGCGCCAGTTCGCCGAGCGCGCCAGGGGCC 530
Db 1131 SerArgArgThrProAsnAsnProGlyAsnPro 1141

RESULT 15

YT35_STRFR
ID YT35_STRFR STANDARD: PRT: 348 AA.
AC P20186;
DT 01-FEB-1991 (Rel. 17, Created)
WT 01-FEB-1991 (Rel. 17, Last sequence update)
16-OCT-2001 (Rel. 40, Last annotation update)
OS Hypothetical 35.5 kDa protein in transposon TN4556.
OC Streptomyces fradiae.
OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID-1906;
RN [1]
RP SEQUENCE FROM N.A.
RC TRANSPOSON-TN4556;
RX MEDLINE-90185236; PubMed-2155856;
RA Siemieniak D.R., Slightom J.L., Chung S.T.;
RT "Nucleotide sequence of Streptomyces fradiae transposable element
TN4556: a class-II transposon related to Tn3.";
RL Gene 86:1-9(1990).
CC -----

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC -----

DR EMBL: M29297; AAA88561.1; -
DR PIR: JQ0431; JQ0431.
KW Hypothetical protein; Transposable element.
SQ SEQUENCE 348 AA; 35520 MW; 3BDF5D83ABBB92A CRC64;

Alignment Scores:

ed. No.:	4.06	Length:	348
Score:	90.00	Matches:	38
Percent Similarity:	43.24%	Conservative:	10
Best Local Similarity:	34.23%	Mismatches:	48
Query Match:	8.43%	Indels:	15
DB:	1	Gaps:	4

US-09-920-953-2 (1-598) x YT35_STRFR (1-348)

QY 116 CTTTCTACGATAGGTGCTGCTGACCGGAGCTGCTCTTCGAGTCCCTGGACA 175
Db 212 ProLeuAlaAlaArgCysSerAlaSerArgAlaGlyCysProThrAlaAlaGlySerLeu 231
QY 176 TGCAGAGCAGACAGCA-----TGAAGCAGGTCAAGTTCATGAGCTTCGTGTTGGCG 226
Db 232 LeuProAlaProArgProProAlaSerAlaSerProGlnAlaAlaProAla 251
QY 227 GAGCAGACCAATCAAGCGCGGAGCATGTACGACGACAGCCCATCTCGTCAAGGGCC 286
Db 252 AlaProSerAlaThrArgLeu-Pro-----ArgArgThrProSerAlaProArgPr 269
QY 287 A-----CGGCCTGGACCGCCACTTTGACAAGATCAAGCAGTACCTTGGAGAC-- 338
Db 269 oSerSerArgProAlaArgProProLleProAlaAlaArgProProArgThrPr 289

QY 339 -----GCTCAAGAGATGGCGCTCAAGCAGGAGTGTGATCCAGCACG 379
Db 289 OGlyThrProArgProAlaAlaAlaArgAlaArgAlaProAlaGlyCysSerProAlaAr 309
QY 380 CCGCCCGAGTGTGGAGTCCACCCCGCGACGA 410
Db 309 gArgThrProSerAlaProThrAspArg 319

Search completed: April 16, 2003, 13:01:03
Job time : 18 secs


```

Query Match: 23.90% Indels: 0
DB: 2 Gaps: 0
US-09-920-953-2 (1-598) x Q8RT58 (1-124)

QY 67 CTGTTTGATGACCTGGCGCGCCAGAGCATGAAGCTGGCGTTGACACCTTCTACGAT 126
DB 4 LeuThrGluLeuGlyGlyAlaAlaAlaValAspLeuAlaValGluLeuPheThrGly 23
QY 127 AAGTGCTGCTGACCGGAGGTGCTGCTTCTTCGAGTCCCTGCACATGCAAGAGCAG 186
DB 24 LysValLeuAlaAspGluArgValAlaAsnArgPhePheValAsnThrAspMetAlaLysGln 43
QY 187 AGATGACAGCGTCAAGTTCATGAGCTTCGTGTTGGCGGAGCAGACCAATACAGGCG 246
DB 44 LysGlnHisGlnLysAspPheMetThrTyrAlaPheGlyGlyThrAspArgPheProGly 63
QY 247 CGAAGCATGACGACGACCCCATCTGCTCAAGGGCCAGCGCTGGACCCGCCAC 306
DB 64 ArgSerMetArgAlaAlaHisGlnAspLeuValGluAlaAsnAlaGlyLeuThrAspValHis 83
QY 307 TTTGACAAGATCAACGATPACCTTGGAGAGACGCTGCAAGAGATGGCGTCAAGCAGGAT 366
DB 84 PheAspAlaIleAlaGluAsnLeuValLeuThrLeuGlnGluLeuAsnValSerGlnAsp 103
QY 367 GTGATCCACGACCGCGCGAGTGTGGAGTCCACCGCGAGCAATTTGACTTNCACAC 426
DB 104 LeuIleAspGluValValThrIleValGlySerValGlnHisArgAsnAspValLeuAsn 123

RESULT 2
Q27224
ID Q27224 PRELIMINARY; PRT: 117 AA.
AC Q27224;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE HEMOGLOBIN
OS Parametrium jenningsi, and
OS Parametrium triaurelia.
OC Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Peniculida;
OC Parametrium.
OX NCBI_TaxID=44029, 44031;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96038820; Pubmed=7578257;
RA Yamauchi K., Tada H., Usuki I.;
RT "Structure and evolution of Parametrium hemoglobin genes.";
RL Biochim. Biophys. Acta 1264:53-62(1995).
DE EMBL: D49689; BAA08540.1;
EMBL: D49688; BAA08539.1;
InterPro: IPR001486; Bac_globin.
DR Pfam: PF01152; Bac_globin; 1.
DR ProDom: PD004840; Bac_globin; 1.
DR PROSITE: PS01213; GLOBIN_FAM.2; 1.
SQ SEQUENCE 117 AA; 12090 MW; 51A8C7F7E1E5106B CRC64;

Alignment Scores:
Pred. No.: 2.34e-05 Length: 117
Score: 149.00 Matches: 31
Percent Similarity: 46.96% Conservative: 23
Best Local Similarity: 26.96% Mismatches: 57
Query Match: 13.96% Indels: 4
DB: 5 Gaps: 1

US-09-920-953-2 (1-598) x Q27224 (1-117)
QY 67 CTGTTTGATGACCTGGCGCGCCAGAGCATGAAGCTGGCGTTGACACCTTCTACGAT 126
DB 3 LeuPheGluLeuGlyGlyAlaAlaAlaValThrAlaValThrGlnPheThrAla 22
QY 127 AAGTGCTGCTGACCGGAGGTGCTGCTTCTTCGAGTCCCTGCACATGCAAGAGCAG 186
DB 23 AsnIleGlnAlaAspAlaThrValAlaAsnPhePheAsnGlyIleAsnMetProThrGln 42
QY 187 AGATGACAGCGTCAAGTTCATGAGCTTCGTGTTGGCGGAGCAGACCAATACAGGCG 246
DB 43 ThrAspLysThrAlaAlaPheLeuCysAlaAlaLeuGlyProAsnAlaThrAlaGly 62
QY 247 CGAAGCATGACGACGACCCCATCTGCTCAAGGGCCAGCGCTGGACCCGCCAC 306
DB 63 ArgAsnLeuLysGluValHisAlaAsnMet-----GlyValHisAlaAlaGln 78
QY 307 TTTGACAAGATCAACGATPACCTTGGAGAGACGCTGCAAGAGATGGCGTCAAGCAGGAT 366
DB 79 PheThrThrValIleGlyHisLeuArgSerAlaLeuThrGlyAlaGlyValAlaAla 98
QY 367 GTGATCCACGACCGCGCGAGTGTGGAGTCCACCGCGAGCAATTTGACTTNCACAC 411

US-09-920-953-2 (1-598) x Q27213 (1-118)
QY 67 CTGTTTGATGACCTGGCGCGCCAGAGCATGAAGCTGGCGTTGACACCTTCTACGAT 126
DB 3 LeuPheGluLeuGlyGlyAlaAlaAlaValThrAlaValThrGlnPheThrAla 22
QY 127 AAGTGCTGCTGACCGGAGGTGCTGCTTCTTCGAGTCCCTGCACATGCAAGAGCAG 186
DB 23 AsnIleAlaAlaAspAlaThrValAlaAsnPhePheAsnGlyIleAsnMetProThrGln 42
QY 187 AGATGACAGCGTCAAGTTCATGAGCTTCGTGTTGGCGGAGCAGACCAATACAGGCG 246
DB 43 ThrAspLysThrAlaAlaPheLeuCysAlaAlaLeuGlyProAsnAlaThrAlaGly 62
QY 247 CGAAGCATGACGACGACCCCATCTGCTCAAGGGCCAGCGCTGGACCCGCCAC 306
DB 63 ArgAsnLeuLysGluValHisAlaAsnMet-----GlyValHisAlaAlaGln 78
QY 307 TTTGACAAGATCAACGATPACCTTGGAGAGACGCTGCAAGAGATGGCGTCAAGCAGGAT 366
DB 79 PheThrThrValIleGlyHisLeuArgSerAlaLeuThrGlyAlaGlyValAlaAla 98
QY 367 GTGATCCACGACCGCGCGAGTGTGGAGTCCACCGCGAGCAATTTGACTTNCACAC 411

```


QY 488 TTTGACGGGGGACCGCAGTTCCGAGCGCGCCAGGGGGCCAGGAGCGCTGCAATCG- 546
 Db 186 exCysAlaGlySerCysGlyAlaArgThrAlaProThrProAlaProThrCysAlaSerP 206
 QY 547 -----TTTGACCGCTTGTCTGC 564
 Db 206 roSerAlaAlaAlaSerSerCysCys 214
 RESULT 6
 ID Q98N74 PRELIMINARY: PRT: 160 AA.
 AC Q98N74: (TREMBlrel. 18, Created)
 DT 01-OCT-2001 (TREMBlrel. 18, Last sequence update)
 DT 01-OCT-2001 (TREMBlrel. 20, Last annotation update)
 DE Hypothetical protein mlr0267.
 GN MLR0267.
 OS Rhizobium loti (Mesorhizobium loti).
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 CC Phyllobacteriaceae; Mesorhizobium.
 NCBI_TaxID=381;
 [1]
 SEQUENCE FROM N.A.
 RP STRAIN-WAFF303099;
 RC MEDLINE-21082930; PubMed-11214968;
 RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
 RA Watanabe A., Idehara K., Ishikawa A., Kawashima K., Kimura T.,
 RA Kishida Y., Miyokawa C., Kohara M., Matsumoto M., Matsuno A.,
 RA Mochizuki Y., Nakayama S., Nakazaki N., Shimo S., Sugimoto M.,
 RA Takeuchi C., Yamada M., Tabata S.
 RT *Complete genome structure of the nitrogen-fixing symbiotic bacterium
 Mesorhizobium loti.;
 RL DNA Res. 7:331-338(2000).
 DR EMBL; AP002994; BAB47888.1; .
 DR InterPro; IPR001486; Bac_globin.
 DR Pfam; PF01152; Bac_globin; 1.
 DR ProDom; PD004840; Bac_globin; 1.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 160 AA; 17777 MW; DBCFCE748FA4EB0B CRC64;

Alignment Scores:
 Pred. No.: 0.0474 Length: 160
 Score: 113.50 Matches: 40
 Percent Similarity: 40.95% Conservativeness: 27
 Best Local Similarity: 24.39% Mismatches: 70
 Query Match: 10.64% Indels: 27
 DB: 16 Gaps: 7

--09-920-953-2 (1-598) x Q98N74 (1-160)

QY 67 CTGTTTGATGACCTGGCGGCGCAGAGGCATGAAGCTGGCGGTGACACCTTCTACGAT 126
 Db 8 LeuTyrgluTrpAlaGlyGlySerAspAlaLeuAsnArgLeuThrGlnThrPheTyrrAsp 27
 QY 127 AAGTGCTGCTGACCGGACCTGCTGCTTCTCGACTCCCTGGACATGCAAGAGCAG 186
 Db 28 LysValAlaLysAspProValGlyProValPheLysAlaMetSer---ProAspHis 46
 QY 187 AAGTGAAGCAGGTCACGTTTCATGCTGCTGTTGGCGGAGCAGACCAATACAAGGC 246
 Db 47 ProSerHisValAlaAlaPheIleGlyGluValPheGlyGlyProLysThrTyrrSerGlu 66
 QY 247 CGAAGCATGTAGAGCGACACGCCCATCTGCTGAAGGGCGGCGCTGGACACCGCCAC 306
 Db 67 Lys-----PheGlyGlyHisArgGluMetValMetHisLysLeuGlyLysHisLeuThr 84
 QY 307 TTGCACAGATCAACCACTAC-----CTTGGAGAGCGCTGCAAGATGGCGTC 357
 Db 85 GluGluGlnArgArgTrpIleAsnLeuLeuAlaAlaAspGluValGlyLeu 104
 QY 358 AAGCAGGATGTGATCCACGACCGGCC-----GGAGTGGTGGAG-----TCC 399
 Db 105 ProAspAspProGluPheArgSerAlaPheMetGlyTyrrValGluTrpGlySerArgLeu 124

QY 400 ACCCGCAGCAATTTGACTTNCACCAACACTGCCACCACTGATTTTTCATTAACCAA 459
 Db 125 AlalysMetAsnSerAsnLeuGlyCluThrCysAspProGlu-----ThrGlu 140
 QY 460 CCCAGCCTGAGCGCTCATTCATCGATTTTGGAGGGGAGCGCCAGTTGCCAGCGCGC 519
 Db 141 PrometProAlaTrpGlyTrp-----GlyVal 149
 QY 520 CCAGGGGGCCCA 531
 Db 150 ProGlyGlyPro 153
 RESULT 7
 ID Q96397 PRELIMINARY: PRT: 640 AA.
 AC Q96397: O04833; O04733;
 DT 01-FEB-1997 (TREMBlrel. 02, Created)
 DT 01-FEB-1997 (TREMBlrel. 02, Last sequence update)
 DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
 DE LRG5.
 GN LRG5.
 OS Chlamydomonas reinhardtii.
 OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
 CC Chlamydomonadaceae; Chlamydomonas.
 OX NCBI_TaxID=3055;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Gloeckner G., Beck C.F.;
 RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Gloeckner G., Beck C.F.;
 RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL; U73817; AAB17561.1; .
 DR EMBL; U73818; AAB39840.1; .
 DR InterPro; IPR001007; VWF_C.
 DR PROSITE; PS01208; VWF_C; UNKNOWN1.
 SQ SEQUENCE 640 AA; 67298 MW; 764EEE7ADC32FC99 CRC64;

Alignment Scores:
 Pred. No.: 0.0885 Length: 640
 Score: 112.00 Matches: 56
 Percent Similarity: 33.94% Conservativeness: 18
 Best Local Similarity: 25.69% Mismatches: 59
 Query Match: 9.98% Indels: 85
 DB: 10 Gaps: 17

US-09-920-953-2 (1-598) x Q96397 (1-640)

QY 590 GCGAAATGGCTGATGCTCTTCAATGCAGCAGGCTGGCAACGATTTTCGACGCTCCT 531
 Db 446 AlaArgValAlaAlaGlyThrSer-----ArgAlaAla 456
 QY 530 GGGCGCCCTGGCGCGCTCGCAACTGGCGCTCCCGCTCAAAATCGATGAATGAGCGC 471
 Db 457 GlyValProArgArg-----LeuGlnArgArgTyrArgArgArg 469
 QY 470 TCAGCTGGGTTGGGTTTAAATGAANAATCAGTTGGTGGCGCAGTTGTTGGGNAAGTCAANT 411
 Db 470 ---GlyArgGlyTrp-----ArgArgArgValArgArgArgGly 482
 QY 410 TCGCGCGGCTGGACTCCACCACCTCCGCGCGCTGGATGCATCTGCTTG-----357
 Db 483 AlaGlyArgAlaValCysThrAlaGlyArgCysCysTrpMetThr---CysLeuPromet 501
 QY 356 -----ACGCCCATC---TCTTGGCAGC 339
 Db 502 TrpGlySerGlyGlyThrTrpTrpArgProLeuMetThrProSerArgThrCysAla 521
 QY 338 GTCTCTCAAGGTACTGCTTGATCTTGTCAAAAGTGG-----CGTGGTCCAGCCCGTGG 285
 Db 522 CysLeuProThrProCys-----CysSerArgTrpLeuArgArgTrpArgCysGlyTrp 539

```
QY 284 ---CCCTTGACAGATGGCGTGTGCTGATACATGCTTCCGGCCCTTGATGCTGCTGCT 228
DB 540 AlaProGlyGlyArgTrpArgCysSer-----LeuCysSerCysTrpArgTrpGlyCys 557
QY 227 CCGCCAAACACGAAGCTCATGAACCTTGACCTGCTTTCATCTTC-----TGCTCT 180
DB 558 SerGlyArgThrProLeu-----LeuProThrTrpValTrpArgArgCysCysArg 575
QY 179 TGCATGTCAGGAGCTCGAAGAGGCGACAGCTCGGGTCCAGCAGCCTTATCGTAG 120
DB 576 Cys-CysArgGlySerArg-----AlaPro----- 583
QY 119 AAGTGTCAACCCCGCTTCATGCTTCTGCGCCGCCAGGTCATCAACACGCTTCTTG 60
DB 584 -ArgCysAsnTrpValCysCysLeu----- 592
QY 59 CGCCCGCATCCCGCTTCCCGCTCTCTGTGCTGCTG-----TGGCCCGCT 14
DB 593 -----ProGlyCysCysCysTrpGlyCysTrpAlaArg 603

RESULT 8
Q9K411
ID Q9K411 PRELIMINARY; PRT; 301 AA.
AC Q9K411;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DE Hypothetical protein SC01322.
GN SC01322 OR 25C61.04C.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae.
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RC STRAIN=A3(2) / M145;
RA Bentley S.D., Chater K.F., Cerdano-Tarraga A.-M., Challis G.L.,
RA Thompson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,
RA Rabinowitz E., Rajandream M.A., Rutherford K., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
RT coelicolor A3(2).";
RL Nature 417:141-147(2002).
JR EMBL: AL359949; CAB95779.1; -.
KW Hypothetical protein.
SQ SEQUENCE 301 AA; 31669 MW; 1322B81BECDD17A6D CRC64;

Alignment Scores:
Pred. No.: 0.241 Length: 301
Score: 106.50 Matches: 44
Percent Similarity: 32.24% Conservative: 15
Best Local Similarity: 24.04% Mismatches: 73
Query Match: 9.49% Indels: 51
DB: 16 Caps: 9

US-09-920-953-2 (1-598) x Q9K411 (1-301)
QY 565 TGCAGCAGGGCTGGCAACGATTTGACGGCTCTTGGGGCCCTGGCGCGCTCGGCAAC 506
DB 107 CysSerSerGlyLeuProValAlaArgHisLeu----- 118
QY 505 TGGCGCTCCCGCTCAAAATCGATGGAACTGAGCGCTCAGCGCTGGGGTTGGTTAATGAAA 446
DB 119 -----LeuThrThr 121
QY 445 ATCAGTTGGGTGGCGCAGCTTGTGG-----GNAAGTCAAAATTCGTCGC 404
DB: 111
```

```
DB 122 ValProTrp-----TrpAspThrValAspLeuLeuAlaHisValValGly 137
QY 403 GGGTGGACCCACCACCTCGGGCGGTGGTGGATCACATCTTGCAGCCCA---TCT 347
DB 138 AlaTrpSerProPro-----ThrAlaAlaSerArgProThrTrpThrProGlySer 154
QY 346 CTTCGACGCTCTCTCCAAAGCTTGTATCTTGTCAAGTGGCGTGTGTCAGGCCGT 287
DB 155 MetThrArgThrAlaGlyTrpSerAlaArgProSerSerThrSerCysGlyThrArgAsn 174
QY 286 GGCCTTGGACCATGGCGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 227
DB 175 AlaProThrProThrGlySerSerAlaThrAlaCysAlaGlyProAlaThrAlaThrSer 194
QY 226 CGCCAAACAGGAGCTCATGAACCTTGACCTGCTTCTCTCTCTCTCTCTCTCTCT 167
DB 195 SerSerAlaArg-----ProSerAlaGlyAlaCysAlaSerThrProGly 209
QY 166 ACTCGAAGAGGCGCAGCAGCTCGGGTTCAGCCA-----GCACCTTATCTAGAAAGTGT 113
DB 210 ProThrArgThrProCysAlaProSerTrpProSerThrAlaProAlaSerArgCys 229
QY 112 CAACCCGACGCTTCATGCTTCTGCGCGCCAGGTCATCAACAGCTTCTTGGCCCGC 53
DB 230 -----ProArgGly-----ArgArgCysGlyProSerAlaArgSerAlaGlyAla 244
QY 52 ATCCCGCTT 44
DB 245 ValProLeu 247

RESULT 9
Q04397
ID Q04397 PRELIMINARY; PRT; 383 AA.
AC Q04397;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DE Hypothetical BWRFl protein.
GN BWRFl OR BGRF2.
OS Epstein-barr virus (strain B95-8) (Human herpesvirus 4).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Gammaherpesvirinae; Lymphocryptovirus.
OX NCBI_TaxID=10377;
RN [1]
RC SEQUENCE FROM N.A.
RA Baer R., Bankier A.T., Biggin M.D., Deininger P.L., Farrell P.J.,
RA Gibson T.J., Hatfull G., Hudson G.S., Satchwell S.C., Seguin C.,
RA Tuffnell P.S., Barrell B.G.;
RT "DNA sequence and expression of the B95-8 Epstein-Barr virus genome.";
RL Nature 310:207-211(1984).
RN [2]
RP SEQUENCE OF 88-145 FROM N.A.
RC STRAIN=B95-8;
RX MEDLINE=88296424; PubMed=28411116;
RA Walls D., Gannon F.;
RT "The expression of novel antigens from the Epstein-Barr virus large
RT internal repeat.";
RL EMBO J. 7:1191-1196(1988).
DR EMBL: V01555; CAA24870.1; -
DR EMBL: V01555; CAA24871.1; -
DR EMBL: V01555; CAA24872.1; -
DR EMBL: V01555; CAA24873.1; -
DR EMBL: V01555; CAA24874.1; -
DR EMBL: V01555; CAA24869.1; -
DR EMBL: V01555; CAA24875.1; -
DR EMBL: X07816; CAA30675.1; -
DR EMBL: V01555; CAA24864.1; -
DR EMBL: V01555; CAA24865.1; -
DR EMBL: V01555; CAA24866.1; -
DR EMBL: V01555; CAA24867.1; -
DR EMBL: V01555; CAA24868.1; -
KW Hypothetical protein.
```


[illegible]

```

Alignment Scores:
Pred. No.:          0.677          Length:          425
Score:             102.00         Matches:         69
Percent Similarity: 27.36%        Conservative:    12
Best Local Similarity: 23.31%     Mismatches:     62
Query Match:       9.09%          Indels:         153
DB:                6              Gaps:           17

US-09-920-953-2 (1-598) x 002661 (1-425)

QY 554 CTGCCAACCATTGTCAGGCTCCTGGG----- 528
   ::  :::::|||||  |||||
Db 97 VAlGlyGlUleuGinProGlyValProPheLeuLeuAspAsnCysSerArGcys 171
   ::  :::::|||||  |||||
QY 527 -----CCCCCTGGGGCGCTCGGCAACTGGCGCTCCCC 495
   |||||  |||||
Db 117 VAlCysGlUlysGlyAlaLeuLeuCysGlUProGly-----GlyCysPro 131
   |||||  |||||
QY 494 GCTCAAAATCGATGGAATGAGCGCTCAGGCTGGGGTTGGGTTAATGAA-----AAT 444
   |||||  |||||  |||||  :::
Db 132 VAlProCysGlyTrpSerAlaTApSerSerTrpClyProCysAspArgSerGlySer 151
   |||||  |||||  |||||  :::
QY 443 CAGTTGGTGGCGGCAAGTTGTTGGGNAAGTCAAT----- 411
   |||||  |||||  |||||
Db 152 GlyLeuAlqAlaArgPheArgSerProSerAsnProProAlaAlaSerGlyAlaPro 171
   |||||  |||||  |||||

```

QY 410 -----TCTCCGCGGTGCAC 396
 Db 172 CysGluGlyGlnArgGlnLeuGlnAlaCysTyrSerAlaCysGlyAlaGluValPro 191
 QY 395 TCCACACCTCCGCGCGGTGCTGGATCACA-----TCTCGCTTGACGCC----- 351
 Db 192 GlyTrpThrProTrpAlaProTrpSerAlaCysSerGlnSerCysLeuValProGlyGly 211
 QY 350 -----ATCTCTTGC----- 342
 Db 212 GlyProAlaLeuArgSerArgSerArgLeuCysProGlyProGlyAspThrSerCysIle 231
 QY 341 -----AGGCTCTCCAGGACTGCTGCTGATCTTGTCAAGTGG 303
 Db 232 GlyGluAlaThrGluGluGluProCysSerProProValCysLeuGlyLeuGlyValTrp 251
 QY 302 CGGTGGTCCAGCGCGTGGCCCTTGACCAGATGGCGTGGTGTGATCATCTCTCGGCC 243
 Db 252 -----GlyGlnTrpAlaAlaTrpSerAlaCysSerAlaPr 263
 QY 242 TTGTATTGCTGCTGCTCCGCCCAACACGAGCTCATGAACCTTGACCTGCTCTCTCTGC 183
 Db 263 oCysAsnGlyGlyVal---GlnThrArg----- 271
 QY 182 TCTTGCATGTCACGGACTCCAAAGAGGCGACGCTCCGGGTGAGCCA-----GC 132
 Db 272 -----GlyArgArgCysSerAlaSerAlaProGlyAspProGlyCysGlnGln 287
 QY 131 ACCTTAT-----CGTAGAGGTGT-----CAACCG----- 107
 Db 287 yProHisSerGlnThrArgaspCysAsnThrGlnProCysThrAlaGlnCysProGlyAs 307
 QY 106 -----CCAGCTTCATG 96
 Db 307 pMetValPheArgSerAlaGluGlnCysArgTrrpGluGlyGlyProCysProGlyLeuCy 327
 QY 95 CCTCTCGCGCGCCAGGTATCA-----ACAGCTTCTTGGCGCCGCAT 51
 Db 327 sLeuAlaArgGlyProGlyValGluCysThrGlyValCysThrAlaGlyCysAla---Cy 346
 QY 50 CCGG-----CTTCCGCGCTCT 35
 Db 346 sProThrGlyLeuPheLeuHisAsnSerSerCysLeuProProSer 361
 RESULT 13
 Q9DWH3 PRELIMINARY; PRT; 629 AA.
 ID Q9DWH3; 01-MAR-2001 (TREMBLrel. 16, Created)
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
 DE Pr5. 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 GN R5.
 OS Rat cytomegalovirus (strain Maastricht).
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
 OC Betaherpesvirinae; MurinegaloVirus.
 OX NCBI_TaxID=79700;
 RN [1]
 RC SEQUENCE FROM N.A.
 RC STRAIN-MAASTRICHT;
 RX MEDLINE-20366325; PubMed-10906222;
 RA Vink C., Beuken E., Bruggeman C.A.;
 RT "Complete DNA sequence of the rat cytomegalovirus genome.";
 RL J. Virol. 74:7656-7665(2000).
 RN [2]
 RC SEQUENCE FROM N.A.
 RC STRAIN-MAASTRICHT;
 RX MEDLINE-20473137; PubMed-11018281;
 RA Gruijthuisen Y.K., Beuken E., Bruggeman C.A., Vink C.;
 RT "Rat cytomegalovirus R89 is a highly conserved gene which expresses a
 RT spliced transcript.";
 RL Virus Res. 69:119-130(2000).
 RN [1]

DR EMBL; AF232689; AAF99116.1. -.
 DR PRINTS; PR01574; TUBBYPROTEIN.
 SQ SEQUENCE 629 AA; 65124 MW; EBAD52655610D8BC CRC64;
 Alignment Scores:
 Pred. No.: 0.738 Length: 629
 Score: 102.00 Matches: 54
 Percent Similarity: 34.60% Conservative: 19
 Best Local Similarity: 25.59% Mismatches: 86
 Query Match: 9.56% Indels: 52
 DB: 12 Gaps: 10
 US-09-920-953-2 (1-598) x Q9DWH3 (1-629)
 QY 26 GCACACAGACAGCGCGGAGTGGCGGCCACAGAACTGTTTGTATGACCTGGCGC 85
 Db 404 AlaProArgArgProAlaGlyProArgGlyAlaSerSerThrAlaProGlyThr 423
 QY 86 GCGCAGAGGACATGAAGCTGGCGGTTGACACCTTCTACGATAAGGTGCTGGCTGCCCGG 145
 Db 424 GlyArgSerAla-----SerProAlaCysTrpThrAlaThr 436
 QY 146 AGCTGCTGC-----CCTCTTCGAGTCCCTGGACATGCCAAGACAGA----- 187
 Db 437 AlaSerCysTrpThrSerProSerSerGlyProThrSerThrAlaThrArgThrArgThr 456
 QY 188 -----ACATGAACGAGCTCAAGTTTCATGAGCTTCGTTGGCGGAGCAG 232
 Db 457 AlaGlyGlyThrGlyArgAlaThrAlaAlaThrSerAspGlyThrAlaAlaSerSerArg 476
 QY 233 ACCAATACAGGCGCGAAGCATGTACGACGACACGCGCCATCTGGTCAAGGGCC----- 286
 Db 477 ThrThrSerProAlaArgProCys-----GlySerProAlaThrSerArgAlaProAla 494
 QY 287 -----ACGGCTCTG-----ACCACCGCCACT 307
 Db 495 SerGlyGlySerAlaArgGlyProThrThrTrpProAlaSerThrAlaAlaSerCysThr 514
 QY 308 TTGACAAAGTCAACGACAGTACCTTGGAGAGACGCTGCAAGAGATGGCGTCAAGCAGGATG 367
 Db 515 ThrArgArgSerSerThr-----CysArgGlyAlaAlaArgProArgAsp 530
 QY 368 TGATCCAGACGCGCGCGGAGTGG-----TGGAGTCCACCGCGCAGCAT 412
 Db 531 GlySerSerProThrProAlaArgSerGlyArgCysProTrpSerProTrpProArg 550
 QY 413 TTGACTTNCACAACTCGCCACCACTGATTTTCATTAACCCACCGCCCTGAGC 472
 Db 551 ProGlySerAlaAlaAlaSerThrGlyArgGlyIleArgGlySerSerProSerSerSer 570
 QY 473 GCTCATTC-----ATCGATTTTGGCGGGGAGCG-----CCAGTT 508
 Db 571 SerArgSerSerAlaThrGlyThrAspSerSerSerGlyThrTrpThrGlyArgProThr 590
 QY 509 GCGCAGCGCGCGCGGCGGCCAGGAGGAGCTGCA 541
 Db 591 GlyThrAlaSerArgSerAlaSerArgProAla 601
 RESULT 14
 Q8SPM4 PRELIMINARY; PRT; 5146 AA.
 ID Q8SPM4; 01-JUN-2002 (TREMBLrel. 21, Created)
 DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
 DE 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 GN SCO-spondin.
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovidae; Bovinae; Bos.
 OX NCBI_TaxID=9913;
 RN [1]


```
OY 410 TCGTCGGGTGACTCCACCACTCGCGGGGGTGTGTGTGATCATCATCTGCTTGACGCC 351
    |||
Db 21 GlnSerProThrSerAsnHisSerProThrSerCys----- 32
OY 350 ATCTCTTGACGGTCTCTCCAAAGTACTGCTTGTATCTTCAAGTGGCGGTGGTCCAGG 291
    |||
Db 33 -----ProProLeCys----- 36
OY 290 CCGTGGCCCTTACCAGATGGCGTGTGCTGCTATCATGCTTCGGCCCTGTATTGGTCT 231
    |||
Db 37 -----ProGlyTyrArgPrpMetCysLeuArgPheLeuPheLeuPhe----- 52
OY 230 GCTCCGCCAAACACGAGCTCATGACCTGACCTGCTTCATCTTC----- 186
    |||
Db 53 -----IleLeuLeuLeuCysLeuIlePheLeuLeuValLeuLeu 65
OY 185 -----TGCTCTTCATGCTCCAGGAGCTCGAAGAGGCGACG 150
    |||
Db 66 AspTyrGlnGlyMetLeuProValCysProLeuIle-ProGlySerSerThrThrSerTh 85
    |||
OY 149 AGCTCCGGGTGACCCAGCACCTTATCGTAGAAGGTGTCAACCCCGCAGCT----- 101
    |||
Db 85 rGlyPro-----CysArgThrCysThrThrProAlaGlnGlyThrSe 99
OY 100 -----TCATGCTTCTCGCGCCGCGGAGGTATCAACAGCTTCTTGGCGCCGAT 51
    |||
Db 99 rMetTyrProSerCysCysThrLysProSerHisGlyAsnCysThrCys-----117
OY 50 CCGCTTCGGCGCTCT-----CTGTGGTGGTCTG 23
    |||
Db 117 eProIleProSerSerTrpAlaPheGlyLysPheLeuTrpGluTrp 132

RESULT 2
US-08-906-865-4
; Sequence 4, Application US/08906865
; Patent No. 6040168
; GENERAL INFORMATION:
; APPLICANT: Greengard, Paul
; APPLICANT: Porton, Barbara
; APPLICANT: Kao, Hung-Teh
; TITLE OF INVENTION: DNA ENCODING THE HUMAN SYNAPSIN III GENE
; TITLE OF INVENTION: AND USES THEREOF
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David A. Jackson, Esq.
; STREET: 411 Hackensack Ave, Continental Plaza, 4th
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE: US/08/906.865
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-202
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-487-5800
; TELEFAX: 201-343-1684
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 696 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
```

```
TOPOLOGY: linear
MOLECULE TYPE: protein
DESCRIPTION: /desc - "Synapsin Ia"
HYPOTHEICAL: NO
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
US-08-906-865-4
```

Alignment Scores:

```
Pred. No.: 0.148 Length: 696
Score: 93.00 Matches: 45
Percent Similarity: 36.97% Conservative: 16
Best Local Similarity: 27.27% Mismatches: 64
Query Match: 8.72% Indels: 41
DB: 3 Gaps: 7
```

US-09-920-953-2 (1-598) x US-08-906-865-4 (1-696)

```
OY 81 GGGCGGCGCAGAGCATGAAGCTGGCGGTGTGACACCTTCTACGATAAGGTGCTGGCTGA 140
    |||
Db 530 GlyArgGlnSerArgProValAlaGlyGly----- 539
OY 141 CCCGAGCTGCTGCCCTTCTTCGAGCTCCCTGACATGCAAGACAGACAAGATCAAGCAGGT 200
    |||
Db 540 ProGlyAlaProProAlaAlaArgProProAlaSerProSerProGlnArgGlnAlaGly 559
OY 201 CAAGTTCATGAGCTTCTGTTGGCGGAGCAGACCAACATACAGGCCCGCAGCATGTACCA 260
    |||
Db 560 ProPro-----GlnAlaThrArgGlnThrSerValSerGlyProAlaProProLys 576
OY 261 CGCACA---CGCCCATCTGTCAAGGGCCAGCGCTGGACACCCGCCACTTTTGACAAGAT 317
    |||
Db 577 AlaSerGlyAlaProProGlyGlnGlnArgGlnGlyProProGlnLysProProGly 596
OY 318 CAAGCAGTACCTTGGAGAGAGCTCCAAAGATGGG----- 353
    |||
Db 597 ProAlaGlyProThrArgGlnAlaSerGlnAlaGlyProValProArgThrGlyProPro 616
OY 354 -----CGTCAAGCAGGATGTGATCCAGCAGC-----CCGCCGAGTGG 391
    |||
Db 617 ThrThrGlnGlnProArgProSerGlyProGlyProAlaGlyAlaProLysProGlnLeu 636
OY 392 TGGAGTCCACCCGCGACGAATTTGACTTNCACCAACTGGG----- 433
    |||
Db 637 AlaGlnLysProSerGlnAspValProProProAlaThrAlaAlaAlaGlyGlyProPro 656
OY 434 CACCACTGATTTTCAT-----TAACCCCAACCCCGAGCTGAGCGCTCAT 478
    |||
Db 657 HisProGlnLeuPheAsnLeuProGluProAlaProProArgProSerLeuSerGlnAsp 676
OY 479 TCCATCGATTTTGAG 493
    |||
Db 677 GluValLysAlaGlu 681
```

RESULT 3

```
US-09-129-668-4
; Sequence 4, Application US/09129668B
; Patent No. 6428010
```

```
; GENERAL INFORMATION:
; APPLICANT: Greengard, Paul
; APPLICANT: Porton, Barbara
; APPLICANT: Kao, Hung-Teh
```

```
; TITLE OF INVENTION: DNA ENCODING THE HUMAN SYNAPSIN III GENE AND USES
; TITLE OF INVENTION: THEREOF
```

```
; FILE REFERENCE: 600-1-202 CIP
; CURRENT APPLICATION NUMBER: US/09/129,668B
```

```
; CURRENT FILING DATE: 1998-08-05
; EARLIER APPLICATION NUMBER: 08/906,865
```

```
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: Patentin ver. 2.0
```

```
; SEQ ID NO 4
; LENGTH: 696
```


TYPE: PRT
ORGANISM: Homo sapiens
US-09-129-668-4

Alignment Scores:
Pred. No.: 0.148 Length: 696
Score: 93.00 Matches: 45
Percent Similarity: 36.97% Conservative: 16
Best Local Similarity: 27.27% Mismatches: 64
Query Match: 8.72% Indels: 41
DB: 4 Gaps: 7

US-09-920-953-2 (1-598) x US-09-129-668-4 (1-696)

```

QY 81 GCGCGCGCAGAGGATGAAGTGGCGGTGACACCTTCTACGATAAGCTGCTGCTGA 140
    |||||: ||| |||||
Db 530 GlyArgGlnSerArgProValAlaGlyGly- 539
QY 141 CCGCGAGCTGCTCCCTTCTTCAGTCCCTGGACATGCAAGAGCAGAGATGAAGCAGGT 200
    |||||: ||| |||||
Db 540 GlyArgGlnSerArgProValAlaArgProProAlaSerProSerProGlnArgGlnAlaGly 559
QY 201 CAAGTTCATGAGCTTCTGTTTGGCGGAGCAGACCAATACAGGCGCGAAGCATGTACGA 260
    |||||: ||| |||||
Db 560 ProPro- -GlnAlaThrArgGlnThrSerValSerGlyProAlaProProLys 576
QY 261 CGGACA- -CGCCATCTGTCAAGGCCACGGCTGGACCGCCACCGCCACTTGCACAGAT 317
    |||||: ||| |||||
Db 577 AlaSerGlyAlaProProGlyGlyGlnArgGlnGlyProProGlnLysProProGly 596
QY 318 CAAGCAGTACCTGGAGAGCGCTGCAAGAGATGG- - 353
    |||||: ||| |||||
Db 597 ProAlaGlyProThrArgGlnAlaSerGlnAlaGlyProValProArgThrGlyProPro 616
QY 354 - - - - -CGTCAAGCAGGATGTATCCAGCAGC- - - - -CGCGCGAGTGG 391
    |||||: ||| |||||
Db 617 ThrThrGlnGlnProArgProSerGlyProGlyProAlaGlyAlaProLysProGlnLeu 636
QY 392 TGGAGTCCACCGCGGAGCAATTCAGCTTCCCAACACTGG- - - - - 433
    |||||: ||| |||||
Db 637 AlaGlnLysProSerGlnAspValProProProAlaThrAlaAlaGlyGlyProPro 656
QY 434 CACCCAACTGATTTTCAT- - - - -TAACCCCAACCCCGCCCTGAGCGCTCAT 478
    |||||: ||| |||||
Db 657 HisProGlnLeuPheAsnLeuProGluProAlaProProArgProSerLeuSerGlnAsp 676
QY 479 TCCATCGATTTTGAG 493
    |||||: ||| |||||
Db 677 GluValLysAlaGlu 681

```

RESULT 4
5198348-1
; Patent No. 5198348
; APPLICANT: Bitter, Grant A.
; TITLE OF INVENTION: EXPRESSION OF EXOGENOUS POLYPEPTIDES
; AND POLYPEPTIDE PRODUCTS INCLUDING HEPATITIS B SURFACE
; ANTIGEN IN YEAST CELLS
; NUMBER OF SEQUENCES: 5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/586,819
; FILING DATE: 24-SEP-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 231,599
; FILING DATE: 08-AUG-1988
; APPLICATION NUMBER: 748,712
; FILING DATE: 26-JUN-1985
; APPLICATION NUMBER: 412,707
; FILING DATE: 30-AUG-1983
; SEQ ID NO: 1
; LENGTH: 226
5198348-1
Alignment Scores:

Pred. No.: 0.11 Length: 226
Score: 92.50 Matches: 43
Percent Similarity: 31.77% Conservative: 18
Best Local Similarity: 22.40% Mismatches: 53
Query Match: 8.24% Indels: 78
DB: 6 Gaps: 11

US-09-920-953-2 (1-598) x 5198348-1 (1-226)

```

QY 512 CCGCAACTGGCGTCCCGCTCAAAATCGATGAAGCTCAGGCTCGGCTGGGTT 453
    |||||: ||| |||||
Db 24 ArgIleLeuThrIlePro- - - - -GlnSerLeuAspSerTrp- - - - -TrpThr 37
QY 452 AATGAAATCAGTGGTGGCGAG- - - - -TTGTTGGNAAGTCAAAATTCGTCGCGGGTG 399
    |||||: ||| |||||
Db 38 SerLeuAsnPheLeuGlyGlySerProValCysLeuGlyGlnAsnSerGlnSerProThr 57
QY 398 GACTCCACCACTCCGCGCGCTGCTGCATCATCTGCTTGACGCCCATCTCTTGCAGC 339
    |||||: ||| |||||
Db 58 SerAsnHisSerProThrSerCys- - - - - 65
QY 338 GTCTCTCAAGGTACTGCTTGATCTTGTCAAAGTGGCGGTGCTCCAGGCGTGGCCCTTG 279
    |||||: ||| |||||
Db 66 - - - - -ProProIleCys- - - - -ProGly 71
QY 278 ACCAGATGGCGGTGCTGCTGCTATCATGCTTCGCCCTTGTATTGGTCTGCTCGCCAAAC 219
    |||||: ||| |||||
Db 72 TyrArgTrpMetCysLeuArgArgPheIleIlePheLeuPhe- - - - - 85
QY 218 ACGAAGTCAATGAACCTGACCTGCTTCATCTTC- - - - - 186
    |||||: ||| |||||
Db 86 - - - - -IleLeuLeuLeuCysLeuIlePheLeuLeuValLeuLeuAspTyrGlnGly 102
QY 185 - - - - -TCCTCTTCATGCTCAGGACTCGAAGAGGCGCAGCTCCGGGTCA 138
    |||||: ||| |||||
Db 103 MetLeuProValCysProLeuIle-ProGlySerThrThrThrThrThrThrThrThrThr 120
QY 137 GCCAGCACCTTATCGTAGAAGGTCAACGCCAGCT- - - - -TC 99
    |||||: ||| |||||
Db 121 - - - - -CysLysThrCysThrThrThrProAlaGlnGlyAsnSerMetPheProSe 136
QY 98 ATGCTTCTCGCGCGCAGGTCATCAACAGCTTCTTTCGCCCGCATCCCGTCCGCC 39
    |||||: ||| |||||
Db 136 rCysCysCysThrLysProThrHisGlyAsnCysThrCys- - - - -IleProIleProSe 154
QY 38 GTCT- - - - -CTGTGCTGCTGG 23
    |||||: ||| |||||
Db 154 rSerTrpAlaPheAlaLysTyrLeuTrpGluTrp 165

```

RESULT 5
US-08-690-473-2
; Sequence 2, Application US/08690473
; Patent No. 5876923
; GENERAL INFORMATION:
; APPLICANT: Leopoldi, Rosario
; APPLICANT: Roizman, Bernard
; TITLE OF INVENTION: HERPES SIMPLEX VIRUS ICP4 AS AN
; TITLE OF INVENTION: INHIBITOR OF APOPTOSIS
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/690,473

ZIP: 77210
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA: US/08/843.659
APPLICATION NUMBER: US/08/843.659
FILING DATE: Concurrently Herewith
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Highlander, Steven L.
REGISTRATION NUMBER: 37,642
REFERENCE/DOCKET NUMBER: ARSB:519
TELECOMMUNICATION INFORMATION:
TELEPHONE: (512) 418-3000
TELEFAX: (512) 474-7577
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1298 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
US-08-843-659-2

Alignment Scores:
Pred. No.: 0.212 Length: 1298
Score: 92.50 Matches: 51
Percent Similarity: 36.59% Conservative: 9
Best Local Similarity: 31.10% Mismatches: 48
Query Match: 8.67% Indels: 56
DB: 4 Gaps: 9

US-09-920-953-2 (1-598) x US-08-843-659-2 (1-1298)

```
Qy 8 CTGCAGCGCGCCACAGCACCACAGAGCGCGGAGCGGATGCGCGCGCAAGAGC 67
Db 156 LeuSerProArgProGlnProProArgAlaGlnProProArgArgHisGly----- 171
Qy 68 TCTTTGATGACCTGGCGCGCAGAGGCATCAAGCTGCGGTTGACACCTCTTACGATA 127
Db 172 -----ArgTProArgProSerAlaSerSer--- 179
Qy 128 AGGTCTGGTACCGGAGCTGCTGCCCTTCTTCGAGTCCCTGACATGCAAGACGAGA 187
Db 180 -----ThrSerSerAspSerGlySerSer-----SerSerSer 191
Qy 188 AGATCAAGCAGCTCAAGTTCATGACCTTCGTTGGCGGAGCAGACCAATACAGGCC 247
> 192 AlaSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSer 211
Qy 248 GAAGCATGTACGACGACACGCCCATCTGGTCAAGGGGCCACGG----- 290
Db 212 Asp-HisAlaArgGluAlaArgAlaValGlyArgGlyProSerSerAlaAlaProAlaAl 231
Qy 291 -----CCTGGACCCCGCCCATCTTGCAGATGCTTGCAGATCAAGAGTACC 328
Db 231 aProGlyArgThrProProProGlyProProProLeu-----SerGluAlaAlaPr 249
Qy 329 TTGG---AGAGACGCTGCAAGAGAT-----GGCGCT----- 356
Db 249 oLysProArgAlaAlaAlaAlaArgThrProAlaAlaSerAlaGlyArgIleGluArgAr 269
Qy 357 -----CAAGCAGGATGTGATCCAGCAGCGCGCGGAGTGGTGGAGTCCAGCC 403
Db 269 gAlaArgAlaAlaValAlaGlyArgAspAlaThrGlyArgPheThrAlaGly---GlnPr 288
Qy 404 GCGACGAATT 413
Db 288 oArgArgVal 291
RESULT 8
US-09-413-814-79
```

Sequence 79, Application US/09413814
Patent No. 6225064
GENERAL INFORMATION:
APPLICANT: Gesellschaft fuer Biotechnologische Forschung mbH
APPLICANT: Bristol-Myers Squibb, Co.
APPLICANT: Bayer, Stefan
APPLICANT: Bloeker, Helmut
APPLICANT: Brandt, Petra
APPLICANT: Cino, Paul M
APPLICANT: Dougherty, Brian A
APPLICANT: Goldberg, Steven L
APPLICANT: Hofle, Gerhard
APPLICANT: Mueller, Joachim
APPLICANT: Reichenbach, Hans
TITLE OF INVENTION: DNA sequences for enzymatic synthesis of polyketide or
FILE REFERENCE: PCT/US 99/23535
CURRENT APPLICATION NUMBER: US/09/413.814
CURRENT FILING DATE: 1999-10-07
EARLIER APPLICATION NUMBER: DE 198 46 493.2
NUMBER OF SEQ ID NOS: 107
SOFTWARE: PatentIn ver. 2.1
SEQ ID NO 79
LENGTH: 1213
TYPE: PRT
ORGANISM: Sorangium cellulosum
US-09-413-814-79

Alignment Scores:
Pred. No.: 0.233 Length: 1213
Score: 92.00 Matches: 54
Percent Similarity: 35.54% Conservative: 5
Best Local Similarity: 32.53% Mismatches: 63
Query Match: 8.62% Indels: 44
DB: 4 Gaps: 11

US-09-920-953-2 (1-598) x US-09-413-814-79 (1-1213)

```
Qy 12 AGACGCGC-----CACCAG-----CACCACAGAGCGCGGAGC 47
Db 151 ArgArgGlyGluHisGlyProArgHisGlnProProLeuArgArgArgProGlyGlyPro 170
Qy 48 GGGATGCGCGGCGCAAGAGCTGTTGTGATGACCT-----GGG 83
Db 171 GlyAlaArgAlaLeuValAlaGluLeuArgProValGlyLeuArgArgValArgAspAla 190
Qy 84 CGGCGCAGAGCATGAAGCTGGCGGTTGACACCTTCTACGATAAGTGCTGCTGACCC 143
Db 191 ArgArgArgArgArgArgArgArgProArgProTyrProArgLeuGlySerGly----- 208
Qy 144 GGAGCTGCTCCCTTCTTTCGAGTCCCTGCACATGCAAGACAGAGAAGTAAAGCA----- 197
Db 209 -----AlaLeuAlaArgAlaArgGly---AlaArgAlaGlyAspArgValGluLeu 224
Qy 198 -----GGTCAAGTTCATGAGTCTGCTGTTGGCGGAGCAGACCA----- 236
Db 225 GlyProGlyAlaAspGlyAspAlaHisGlyArgValProArgArgGlyArgProGlyAla 244
Qy 237 ---ATACAGGCGCGCAAGCATGTACGACGACACACCCCATCTGCTGAGGGCCAGCGCT 293
Db 245 ValValAlaProProProHisAspGluArgArgLeuAspProAlaGluAlaProArgSer 264
Qy 294 GGACCAAGCCCACTTTGACAGATCAAGCAGTACCTTGGAGAGAGAGCTCCAAAGAGATCGG 353
Db 265 HisProArgGlyLeu-----ProArgAlaProArgArgGluProArgArg----- 279
Qy 354 CGTCAAGCAGGATGTGAT-----CCAGCAGCGCGCGGAGTGGTGGAGTGC 398
Db 280 ArgAspArgGlyValAspLeuValAspArgProProAspArgGlyArgArgProGlyVal 299
Qy 399 -----CACCAGCGAGCA 410
||||| |||
```


Alignment Scores:

Pred. No.: 0.489 Length: 882
Score: 88.50 Matches: 68
Percent Similarity: 33.73% Conservative: 16
Best Local Similarity: 27.31% Mismatches: 63
Query Match: 8.29% Indels: 102
DB: 4 Gaps: 14

US-09-920-953-2 (1-598) x US-09-413-814-78 (1-882)

QY 20 CCACACACACACAG-----AGACGGCGGAAGCGGATCGCGCGCAAGAGCTGTTG 73
Db 590 ProGluAlaProArgAlaArgHisArgAlaArgAlaProArgValArgArg 609
QY 74 ATGACCTGGCGGCGCC---AGAAGCGATGAAGCTGGCGGTGCACCTTACCATAGG 130
Db 610 LeuVal-GlyArgArgLeuArgArgAlaArgAlaLeuArgArgLeuArgAlaG 629
QY 131 T-----GCTGCTGACCGCGAGCT----- 149
b 629 yProAlaPheProAlaAlaGlyAlaProGlyAlaValArgArgLeuArgArgSerProAl 649
QY 150 ----GCTGCCCTTCCGATCCCTGGACATGCAAGAGCA-----GAAGATGAAGC 196
Db 649 aGlyValAlaValArgArgGlyProGly---GlyArgAlaProValLeuAspGluAl 668
QY 197 A----- 197
Db 668 aLeuGlyArgAlaAlaProGlyProArgAlaGlyArgProArgAlaGlyGlyAlaLysAs 688
QY 198 -----GGTCAAGTTCATGAG----- 212
Db 688 pValAlaArgGlyAlaAlaGluProArgArgArgGlyArgAlaHisProProAspGlnG 708
QY 213 ----CTTCGTGTTGGCGGAGCAGACCAATCAAGGCGCGGATGACGACGACGACG 268
Db 708 yValLeuArgAlaGlyGlyAspLeuProLeuHisGlyProAlaArgArgValGlnAlaAr 728
QY 269 CCCATCT-----GGTCAAGGCGGCGGCGCTGGA-----CCACC 301
Db 728 gProAlaProAlaHisArgAlaArgGlyProArgArgGlyHisArgArgArgGluProG 748
QY 302 GCCA-----CTTTGACAGATCAAGCAGTACCTTGAG 334
Db 748 uProArgArgAspArgAlaAspArgLeuLeuArgGlnProAlaArgAlaProAspAr 768
QY 335 ACACCGCTGCAAGA---GATGGCGCTCAAGCAGGATGTATCCAGCAGCGCGCGAGTGG 391
b 768 gLeuArgGlyArgProAspValArgArgAlaGlyProProArgAlaArgArgAlaAr 788
QY 392 TGGAGTCCACCGCGGAGCAATTTGACTTNCACCAACTGCGCACCACTGATTTTCAT 451
Db 788 gGlyValArg----- 791
QY 452 TAACCAACACCGCGCTGAGCGCTCATTTCCATCGATTTTGAGCGGGAGCGCCAGTTGCC 511
Db 792 -----AlaProGlyProAlaValArgProAla-----ArgArgGlyPr 804
QY 512 GAGCGCGCGCGGCGCGGCGGAGC 536
Db 804 oAlaAlaGluGlyGlyArgArgAla 812

RESULT 15
PCT-US96-10602-6

; Sequence 6, Application PC/TUS9610602
; GENERAL INFORMATION:
; APPLICANT: The General Hospital Corporation
; TITLE OF INVENTION: INHIBITION OF HEPATITIS B REPLICATION
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston

STATE: MA
COUNTRY: USA
Zip: 02110-2804
COMPUTER READABLE FORM: disk
MEDIUM TYPE: Floppy
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US96/10602
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA: 60/017,814
APPLICATION NUMBER: 60/017,814
FILING DATE: 20-JUN-1995
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Clark, Paul T.
REGISTRATION NUMBER: 30,162
REFERENCE/DOCKET NUMBER: 00786/282001
TELEPHONE: 617/542-5070
TELEFAX: 617/542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 397 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US96-10602-6

Alignment Scores:
Pred. No.: 0.409 Length: 397
Score: 88.00 Matches: 49
Percent Similarity: 30.90% Conservative: 23
Best Local Similarity: 21.03% Mismatches: 62
Query Match: 7.84% Indels: 99
DB: 5 Gaps: 13

US-09-920-953-2 (1-598) x PCT-US96-10602-6 (1-397)

QY 563 CAGCAAGCGGTGGCAACGATTTTCAGGCTCCTGGGCGGCGGCTCGGCAACTG 504
Db 151 ArgArgGlyArgSerProArgArgArgThrProSerPro-----ArgArgArgArgSer 168
QY 503 CGGCTCCCGCTCAAAATCGATGGAATCAGCGCTCAGCG----- 465
Db 169 GlnSerProArgArgArgArgSerGlnSerArgLeuGlyProLeuLeuValLeuGlnAla 188
QY 464 -----TGGGTTGG 456
Db 189 GlyPhePheLeuLeuThrArgIleLeuThrIleProGlnSerLeuAspSerTrp---Trp 207
QY 455 GTTAATGAAATACATGTTGGGTGGCGAG-----TTGTTGGGNAAGTCAAAATTCGTCGCGG 402
Db 208 ThrSerLeuAsnPheLeuGlyGlyThrValCysLeuGlyGlnAsnSerGlnSerPro 227
QY 401 GTGACTCCACCACTCCGCGGCGGCTGCTGGATCATCATCTCTTACGCCCATCTCTTGC 342
Db 228 ThrSerAsnHisSerProThrSerCys-----ProProThrCys 240
QY 341 AGCTCTCTCAAGGTACTGCTTGTATCTTGTCAAGTGGCGGTGCTCCAGCGCGTGGCCCC 282
Db 241 -----ProGlyTyr----- 243
QY 281 TTGACCATGAGGTGGGTGTGGCTGTACATGCTTGGCGCCTTGTATTGCTGTCTCCGCCA 222
Db 244 -----ArgTrpMetCysLeuArgArgPheIlePheLeuPhe----- 256
QY 221 AACACGAAGCTCATGAACCTTGACCTGCTTTCATCTTC----- 186
Db 257 -----IleLeuLeuLeuCysLeuIlePheLeuValLeuLeuAspTyrGln 272

```

QY 185 -----TGCTCTTGCATGTCAGGACTCGAAGAGGCGCAGAGTCGCGG 141
Db 273 GlyMetLeuProValCysProLeuIle-ProGlySerSerThrThrSerThrGlyPro-- 291
QY 140 TCAGCCAGCACCTTATCGTAGAGGTGTCAACGCCAGCT----- 101
Db 292 -----CysArgThrCysMetThrThrAlaGlnGlyThrSerMetTyrPr 306
QY 100 -TCATGCCCTTCTGCGCGCCAGGTCAACACAGCTTCTTGGCCCGCATCCCGTTCC 42
Db 306 oSerCysCysCysThrLysProSerAspGlyAsnCysThrCys-----IleProllePr 324
QY 41 GCCGTCT-----CTGTGGTGCTGG 23
Db 324 oSerSerTrpAlaPheGlyLysPheLeuTrpGluTrp 336

```

Search completed: April 16, 2003, 13:02:35
 ~b time : 20.5 secs

US-09-920-953-2 (1-598) x US-09-821-877-2 (1-389)

```
OY 581 GCTGATGCTCTTCAATCAGCAAGGGCTGGCAACGATTTGACGGCTCTCGGGCCCT 522
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 132 AlaGlySerSer-----SetGlyThrValAsnProValProThrThrValSerPro 149
OY 521 GGGCCGCT-----CGCAACTGGCGCTCCCGCTCAAAATCGAATGAGCGTCA 468
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 150 IleSerSerIlePheSerArgIleGlyAspProAlaArgAsnMetGluAsnIleThrSer 169
OY 467 GGC-----465
Db 170 GlyPheLeuGlyProLeuLeuValLeuGlnAlaGlyPhePheLeuLeuThrArgIleLeu 189
OY 464 -----TGGGGTTGGGTTAATGAAATCAGTTGGTGGCGCGCAG 429
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 190 ThrIleProGlnSerLeuAspSerTrp---TrpThrSerLeuAsnPheLeuGlyGlyThr 208
OY 428 -----TTGTTGGGNAAGTCAAAATTCGTCGGCGGTGGACTCCACACTCCCGCGCGTGC 375
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 209 ThrValCysLeuGlyGlnAsnSerGlnSerProThrSerAsnHisSerProThrSerCys 228
OY 374 TGGATCACATCTCTTGCACGCCCATCTCTTGCAGCGTCTCTCCAAAGGTACTGCTTGATC 315
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 229 -----ProProThrCys-----ProGlyTyr-----235
OY 314 TTGTCAAAGTGGCGGTGTCCAGCGCTGGCCCTTGACGAGTGGCGGTGTGCGTGTAC 255
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 236 -----ArgTrpMetCysLeuArgArg 242
OY 254 ATGCTTCGGCCCTTGATTTGCTGCTGCCGCCAAACGAGTCACTGACCTGACCTGC 195
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 243 PheIleIlePheLeuPhe-----IleLeuLeuLeuCys 253
OY 194 TTCATCTC-----TGTCTTGTGATG 174
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 254 LeuIlePheLeuLeuValLeuLeuAspTyrGlnGlyMetLeuProValCysProLeuIle 273
OY 173 TCCAGGAGCTCGAAGAGGGCAGCGCTCGGGTCCAGCCAGCACCTTATCTAGAAGTG 114
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 274 -ProGlySerSerThrThrSerThrGlyPro-----CysArgAlaCys 287
OY 113 TCACCCCGCAGCT-----TCATGCTTCTGCGCGCCCGCCAGGTCA 75
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 287 sThrThrProAlaGlnGlyThrSerMetTyrProSerCysCysThrLysProSerAs 307
OY 74 TCAACAGCTTCTTGGCGCCGATCCCGCTTCCCGCTCT-----CT 33
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 307 pGlyAsnCysThrCys-----IleProIleProSerSerTrpAlaPheGlyLysPheLe 325
OY 32 GTGCTGCTGG 23
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 325 uTrpGluTrp 328
```

RESULT 2

```
US-09-825-288A-2
; Sequence 2, Application US/09825288A
; Publication No. US20020192822A1
; GENERAL INFORMATION:
; APPLICANT: LEOPARDI, ROSARIO
; APPLICANT: ROIZMAN, BERNARD
; TITLE OF INVENTION: HERPES SIMPLEX VIRUS ICP4 IS AN INHIBITOR OF APOPTOSIS
; FILE REFERENCE: ARCD:317USC1
; CURRENT FILING DATE: 2001-04-02
; PRIOR FILING DATE: 2001-04-02
; PRIOR APPLICATION NUMBER: 09/259,821
; PRIOR FILING DATE: 1999-03-01
; PRIOR APPLICATION NUMBER: 08/690,473
; PRIOR FILING DATE: 1996-07-26
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
```

LENGTH: 1298

TYPE: PRT

ORGANISM: HERPES VIRUS, TYPE 1

US-09-825-288A-2

Alignment Scores:

```
Pred. No.: 1.97 Length: 1298
Score: 92.50 Matches: 51
Percent Similarity: 36.59% Conservative: 9
Best Local Similarity: 31.10% Mismatches: 48
Query Match: 8.67% Indels: 56
DB: 9 Gaps: 9
```

US-09-920-953-2 (1-598) x US-09-825-288A-2 (1-1298)

```
OY 8 CTGCAGACGCGGCCACACAGACGCGGAGCGGATCGGGCGCAAGANGC 67
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 156 LeuSerProArgProProAlaGlnProProArgArgArgArgHisGly-----171
OY 68 TGGTTGATGACCTGGCGGCCGAGAGGCATGAAGCTGGCGTTTCACACCTTCTACGATA 127
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 172 -----ArgTrpArgProSerAlaSerSer---179
OY 128 AGGTGCTGGCTACCCGAGCTGCTGCCCTTCTTCGAGTCCCTCGACATGCAAGACGAGA 187
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 180 -----ThrSerSerAspSerGlySerSerSer-----SerSerSer 191
OY 188 AGATGAAGCAGGTCAAGTTGATGAGCTTCGTTGGCGGAGCAGACCAATACAGGGCC 247
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 192 AlaSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSer 211
OY 248 GAACCATCTACGACGCGCACGCGCCCTCTGTCTCAAGGGCCACCGG-----290
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 212 Asp-HisAlaArgGluAlaArgAlaValGlyArgGlyProSerSerAlaAlaProAlaAl 231
OY 291 -----CCTGCACACCCCTCTTTCACAAGATCAAGACGATGACGATGAC 328
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 231 aProGlyArgThrProProProGlyProProLeu-----SerGluAlaAlaPr 249
OY 329 TTGG---AGAGACGCTCAAGAGAT-----GGGCGT-----356
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 249 OlySerProArgAlaAlaAlaArgThrProAlaAlaSerAlaGlyArgIleGluArgArg 269
OY 357 -----CAAGCAGGATGTATCCAGCAGCGCGCGGAGTGGTGGAGTCCACCC 403
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 269 gAlaArgAlaAlaValAlaGlyArgAspAlaThrGlyArgPheThrAlaGly---GlnPr 288
OY 404 GCGACGAATT 413
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 288 oArgArgVal 291
```

RESULT 3

```
US-09-903-456-77
; Sequence 77, Application US/09903456
; Patent No. US20020138874A1
; GENERAL INFORMATION:
; APPLICANT: Abbott Laboratories
; APPLICANT: Mukerji, Pradip
; APPLICANT: Leonard, Amanda Eun-Yeong
; APPLICANT: Huang, Yung-Sheng
; APPLICANT: Pereira, Suzette L.
; TITLE OF INVENTION: ELONGASE GENES AND USES THEREOF
; FILE REFERENCE: 6407.US.P3
; CURRENT FILING DATE: 2001-07-11
; PRIOR FILING DATE: 2001-07-11
; PRIOR APPLICATION NUMBER: US 09/624,670
; PRIOR FILING DATE: 2000-07-24
; PRIOR APPLICATION NUMBER: US 09/379,095
; PRIOR FILING DATE: 1999-08-23
; PRIOR APPLICATION NUMBER: US 09/145,828
; PRIOR FILING DATE: 1998-09-02
; NUMBER OF SEQ ID NOS: 116
; SOFTWARE: FastSeq for Windows Version 4.0
```



```

Alignment Scores:
Pred. No.: 2.26 Length: 175
Score: 89.00 Matches: 50
Percent Similarity: 36.36% Conservatives: 22
Best Local Similarity: 25.25% Mismatches: 74
Query Match: 7.93% Indels: 52
DB: Gaps: 11

US-09-920-953-2 (1-598) x US-10-012-542-472 (1-175)
QY 533 CTTGGCCCGCTGGCGCTCGCAACTGGCGCTCCCGCTCAAAATCATGGAATGAG 474
Db 9 ProGlyLysProThrArg---ArgGlyLeuGlyLeuProThrAlaSer----- 23
QY 473 CGCTCAGCTGGGGTTGGTGTAAATGAAATCACTTGGTGGCGCAGTTGTTGGGNAAGTCA 414
Db 24 ---SerGlyTrpValTrp----- 28
" 413 AATTCTCGCGGGTGGCTCCACCCTCGCGCGCGCTGGTGGATCATCATCTGCTGTGAGC 354
Db 29 ArgArgCysIleAlaSerTrpGlyThrAlaThrAlaAlaTrpProCysSerCysGlyThr 48
QY 353 CCATC-----TCTTGCAGCGTCTCTCCAAAGTACTGCTGTGATCTGTCAAAGTGG 303
Db 49 GlyMetAlaThrProSerCysSerPro-----CysThrTrpValAlaArgThr 66
QY 302 CGGTGTCCAGGCGGTGGCCCTTGACAGATGGCGTGGCTGCTGCTGACATGCTTCCGGCCC 243
Db 67 Arg-----ProIle-----AlaCysSerSerLeuHisProTrpPro 78
QY 242 TWTATTGTCTGCTCCGCCCAACACGACGCTCATGAAC-----TGGACC 198
Db 79 AlaSerTrpAlaProProSerHisProAlaAlaSerProTyProSerProLeuGly 98
QY 197 TGTTCATCTTCTGCTTTCATGTCAGGAGTCCGAAGAGGAGC---AGCTCCGGG 141
Db 99 ThrArgIleThrSerAlaGlyThrArgThrAlaProArgAlaSerLeuGluAlaGly 118
QY 140 TCAGCCAGCACCTTATCGTAGAGGTGTCAACCGCAGCTTCTGCTGCGCGGCC 81
Db 119 GlyLeuAlaProAlaAlaIleProThrPheAsnGlyProValLeuPro---AlaProSer 137
QY 80 AGCTCATCAACAGCTCTTGGCC----- 57
Db 138 HisSerSerGlyArgSerLeuArgGluSerSerGlyArgProAlaGlyArgTyTrp 157
QY 56 CCGCATCCCGCTTCCCGCTCTCTGTTGCTGGTGGCGCGCTCTGCGAGGACC 3
Db 158 ProLeuGlnAlaThrThrMetLeuIleGlnProMetAlaAlaGluAlaAlaSer 175

RESULT 9
US-10-102-806-543
; Sequence 543, Application US/10102806
; Publication No. US20030054421A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA103P1
; CURRENT APPLICATION NUMBER: US/10/102,806
; CURRENT FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: 09/925,298
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05881
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 846
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 543
; LENGTH: 352
; TYPE: PRT
; ORGANISM: Homo sapiens

FEATURE:
; NAME/KEY: SITE
; LOCATION: (5)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (154)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (167)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-10-102-806-543

Alignment Scores:
Pred. No.: 3.54 Length: 352
Score: 88.00 Matches: 54
Percent Similarity: 35.93% Conservatives: 29
Best Local Similarity: 23.38% Mismatches: 80
Query Match: 8.25% Indels: 68
DB: Gaps: 11

US-09-920-953-2 (1-598) x US-10-102-806-543 (1-352)
QY 14 AGCGGCCACCACGCA-----CCACAGACGGCGGAAGGGGATCGGGCGCAAGA 64
Db 10 ThrArgProMetAlaAlaGluProGlnGlnLysGlnGluProLeuGlySerAsp 29
QY 65 AGCTGTTTGTATGACC-----TGGCGCGCGCAGAGGCATGA 100
Db 30 SerGluValLeuThrValTrpProMetLysProSerTrpLeuSerArgThrGluPhe 49
QY 101 AGCTGGCGGTTGACACCTTCTACGATAAGTGC-----TGGCTGACCGGAGCTGC 151
Db 50 SerLysArgLeuLeuCysArgThrLeuTrpCysGlnSerGlyTrpSerSerArgSerTy 69
QY 152 TGCCCTTCTTCAGTCCCTGGACATGCAAGACGACAGATGAAGCAGGTCAAGTTCATGA 211
Db 70 ThrArgSerMetLeuLysMetThrThrSerIleAsnArgArgSerArgThrSerThr 88
QY 212 GCTTCGTGTTGGCGGAGCAGCAATACAAGGGCGGAGCATGTACGACGACACGCC 271
Db 89 -----LysSerThrArgThrSerAlaArgProGlyLeuThr----- 100
QY 272 ATCTGTCAAGGGCCACGCGCTGGACCGCCACTTTCACAGATCAAGACGATACCTTG 331
Db 101 -----AlaThrValSerIleGlyLeuSerAspSerProThr 112
QY 332 GAGACACGCTGCAAGAGATGGCGTCAAGCAGGATGTATCCAGCACGCCGCCGAGTGG 391
Db 113 TrpArgHisCys-----TrpMetThrAlaArgSerCysSerGlyCluLysGlyGlyHis 130
QY 392 TGGAGTCCACCGCGCAGCAATTG-----ACT 418
Db 131 TrpAlaProArgGlnValGlyValTrpLeuLeuProGlyArgValGlyCysValSerSer 150
QY 419 TNCCCAACAACCTGCGCACCCCACTGATTTTCATTAAACCCACCCAGCTGAGCGCTCAT 478
Db 151 ArgValSer---SerPheProGlyAsp----- 159
QY 479 TCATCATGAT-----TTTGAGCGGGAGCGCCAGTTCGCGACGCGCCAGGGGCC 529
Db 160 GlyLeuAspSerGlyLeuAla---ArgGlySerAlaValSerAlaLeuAlaSerGlyLeu 179
QY 530 CAGGAG-----CCTGCAATCTGTTTCCAGCCCTTCTGCT 562
Db 180 ValGluGluProMetLeuGlyProProPheHisProThrProArgPheLysAlaValSer 199
QY 563 GCATTGAAGACCATCAGCCATTTTCGCCACCAA 595
Db 200 Ala---LysSerLysGluAspLeuValSerGln 209

RESULT 10
US-09-812-862-6
; Sequence 6, Application US/09812862

```

Db	228	ThrSerAsnHisSerProThrSerCys	-----ProProThrCys	240
Oy	341	AGCGTCTCTCCAAGTACTGCTTGATCTTGTCAAAGTGGCGGTGGTCTCAGCGCGTGGCC	282	
Db	241	-----ProGlyTyr	-----	243
Oy	281	TTGACAGAGATGGCGGTGGTGGTGTACATGCTTTCGGCCCTTGTATTGGTGTCTCGGCCA	222	
Db	244	-----ArgTrpMetCysLeuArgArpHeIleIlePheLeuPhe	-----	256
Oy	221	AACACGAAGCTTGAACCTTGATGCTCTCATCTTC	-----	186
Db	257	-----IleLeuLeuCysLeuIlePheLeuValLeuLeuAspTyrGln	272	
Oy	185	-----TGCTCTTCATGTCAGGGACTCAAGAAGGCCAGCAGCTCCGGG	141	
Db	273	GlyMetLeuProValCysProLeuIle-ProGlySerSerThrThrSerThrGlyPro	291	
Oy	140	TCAGCCAGCACCTTATCGTAGAGGTGTCAACGCCAGCT	101	
Db	292	-----CysArgThrCysMetThrThrAlaGlnGlyThrSerMetTyrPr	306	
Oy	100	-TCATGCGCTTCTCGCGCCGCCAGGTGCATCAACAGCTTCTTGGCGCCGCATCCCGCTTCC	42	
Db	306	oSerCysCysCysThrLysProSerAspGlyAsnCysThrCys	-----IleProIlePr	324
Oy	41	GCGGTCTN-----CTGTGGTGGTGG	23	
Db	324	oSerSerTrpAlaPheGlyLysPheLeuTrpGluTrp	336	
RESULT 11				
US-10-213-509-5				
; Sequence 5, Application US/10213509				
; Publication No. US20030054485A1				
; GENERAL INFORMATION:				
; APPLICANT: Welss, Joseph				
; APPLICANT: Scott, Matthew				
; TITLE OF INVENTION: JELLY BELLY GENES AND THEIR USES				
; FILE REFERENCE: STAN-232				
; CURRENT APPLICATION NUMBER: US/10/213,509				
; CURRENT FILING DATE: 2002-08-06				
; PRIOR APPLICATION NUMBER: 60/311,720				
; PRIOR FILING DATE: 2001-08-09				
; NUMBER OF SEQ ID NOS: 5				
; SOFTWARE: FastSeq for Windows Version 4.0				
; SEQ ID NO 5				
; LENGTH: 4123				
; TYPE: PRT				
; ORGANISM: H. sapiens				
US-10-213-509-5				
Alignment Scores:				
Pred. No.: 7.85 Length: 4123				
Score: 88.00 Matches: 60				
Percent Similarity: 28.35% Conservative: 14				
Best Local Similarity: 22.99% Mismatches: 76				
Query Match: 7.84% Indels: 111				
DB: 9 Gaps: 15				
US-09-920-953-2 (1-598) x US-10-213-509-5 (1-4123)				
Oy	578	GATGGCTCTTCAATGCACGAAGGCTGGCAACAGATTGTGCAGGCTCTCT	-----	531
Db	1427	AspGlySerAspGluGlyProAspAlaCysValGluAlaProAlaProAlaMetArg	144	
Oy	530	GGGCGCCCTGGG-----CGCGCTCGGCAACTGGCGCTCCCG	495	
Db	1447	GlyProProGlyGlnAlaGlyGlyProThrSerSerArgAlaProSerProPro	146	
Oy	494	GCTCAAAATCATGGAATAGCCGCTCAGGCTGGGGTGGTGTATGAANAATCAGTTGGGT	435	
Db	1467	ProGluAlaGlnGlyGluGlyArgLysGly	-----	147

OY 434 CCGCAGTTGTTGGGNAAGTCAAAATTCCTCGCGGGTGACATCCACCACTCGGGG---GCC 378
Db 1477 -----GlnGluArgSerArgThrHisLeuThrValProAlaGlySer 1490
OY 377 TGTGGATACATCCCTGCTTACAGCCCATCTCTTGACGCTCTCTCAAGGTACTGCTTG 318
Db 1491 ThrGlnLeuProLeuCysProGlyLeuPheProCysGlyValAlaProGlyLeuCysLeu 1510
OY 317 ATCT-----TGTCNAAGTGGC----- 302
Db 1511 -ThrProGluGlnLeuCysAspGlyIleProAspCysProGlnGlyGluAspGluLeuAs 1530
OY 301 -----GCTGTCAGCCCGTGGC-----CCITGACCATGCG 270
Db 1530 pCysGlyGlyLeuProAlaLeuGlyGlyProAsnArgThrGlyLeuProCysProGlu-- 1549
OY 269 GCGTGTGCTGTACATGC---TTCGGCCCTTGTATTGTCTGCTCGCCGCAACACGAAG 213
Db 1550 -----TyrThrCysProAsnGlyThrCysIleGlyPhe----- 1560
OY 212 CTCATGAAGTTCACCTGCTTCATCTCTGCTCTTGTGATGTCAGGACTCGAAGAGGCG 153
Db 1561 -----GlnLeuValCysAspGly---GlnProAspCys 1570
OY 152 ACCAGCTCCGGGTACG-----CCAGCACCTTATCGT----- 122
Db 1570 sGlyArgProGlyGlnValGlyProSerProGluGlnGlyCysGlyAlaTrpGlyPr 1590
OY 121 -----AG 120
Db 1590 oTrpSerProTrpGlyProCysSerArgThrCysGlyProTrpGlyGlnGlyArgSerAr 1610
OY 119 AAGGTGTCAACCGCCAGCTTCATGCTCTGCGCCGC-----CCAGGTTCATCAAC 69
Db 1610 gArgCysSerProLeuGlyLeuLeuValLeuGlnAsnCysProGlyProGluHisGlnSe 1630
OY 68 ACCTTCTTGGCGCCGATCCCGCTTCCGCG-----TCTCTGTGTGTGTGTGTG 21
Db 1630 rGlnAlaCysPheThrAlaAlaCysProValAspGlyGluTrpSerThrTrpSerProTr 1650
OY 20 G 20
Db 1650 p 1650

RESULT 12
US-09-956-425-4
; Sequence 4, Application US/09956425
; Patent No. US20020045192A1

GENERAL INFORMATION:
; APPLICANT: Kriwacki, Richard
; APPLICANT: Bothner, Brian
; APPLICANT: Lewis, William
; TITLE OF INVENTION: Arf and Hdm2 Interaction Domains and Method of Use Thereof
; FILE REFERENCE: 1340/1/035
; CURRENT APPLICATION NUMBER: US/09/956,425
; CURRENT FILING DATE: 2001-09-19
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 132
; TYPE: PRT
; ORGANISM: Homo sapiens

US-09-956-425-4
Alignment Scores:
Pred. No.: 2.89 Length: 132
Score: 87.50 Matches: 50
Percent Similarity: 35.48% Conservative: 16
Best Local Similarity: 26.88% Mismatches: 53
Query Match: 67
Indels: 67
Gaps: 10
DB: 10

US-09-920-953-2 (1-598) x US-09-956-425-4 (1-132)

OY 595 TTGTTGGAAAAATGGCTGATGGCTCTTAATGCAGCAAGGCTGGCAAAACGATTTCCAGG 536
Db 1 MetValArgArgPheLeuValThrLeuArg-----IleArgArg 13
OY 535 CTCCTGGGCCCCCTGGCGCG-----CTCGGCAAC 506
Db 14 AlaCysGlyProProArgValArgValPheValValHisIleProArgLeuThrGlyGlu 33
OY 505 TGGCGCTCCCGCTCAAAATCGATGGAATGAGCGCTCAGCGTGGGTGGTTAATGAA 446
Db 34 TrpAlaAlaPro-----GlyAlaProAlaAlaValAlaLeuVal--- 46
OY 445 ATCAGTTGGTGGCGCAGTTGTTGGGNAAGTCAAAATTCCTCGCGGGTGGACTCCACCAC 386
Db 47 LeuMetLeuLeuArgSer-----GlnArgLeuGlyGlnGlnProLeu 60
OY 385 CGCGCGCGTCTGGATCACATCTGCTTGACGCCCATCTCTTGCACCGTCTCTCCCAAGCT 326
Db 61 ProArgArgProGlyHisAspAspGlyGlnArgProSerGlyGlyAlaAla--- 77
OY 325 ACTGCTTGATCTGTCAAAGTGGCGTGTCCAGGCCGCTTGCACGATGGCGCT 266
Db 78 -----AlaAlaProArgArgGly----- 83
OY 265 GTGCGTGTACATGCTTCGGCCCTTGTATTGTCTGCTCCGCCCAACACGAAAGCTATGA 206
Db 84 -----AlaGlnLeuArgArgProArgHisSer--- 92
OY 205 ACTTGACCTGTCTATCTTCTGCTCTGTCATGCTCCAGG-----ACTCGAAGAGG 155
Db 93 -----HisProThrArgAlaArgArgCysProGlyGlyLeuProGlyHisAlaGly 109
OY 154 GCAGCAGCTCCGGTTCAGCCAGCCTTATCGTAGAAGGTGTCAACGCCAGCTTCATGC 95
Db 110 GlyAlaAlaProGlyArgGlyAlaAlaGlyArgAlaArgCysLeuGlyProSerAla--- 128
OY 94 CTTCTGCGCCGCCAGGT 77
Db 129 -----ArgGlyProGly 132

RESULT 13

US-09-220-920-116
; Sequence 116, Application US/09220920
; Patent No. US20020002269A1
GENERAL INFORMATION:
; APPLICANT: Milbrandt, Jeffrey D.
; APPLICANT: Baloh, Robert H.
; TITLE OF INVENTION: Artemin, A No. US20020002269A1e1 Neurotrophic Factor
; FILE REFERENCE: 6029-7996
; CURRENT APPLICATION NUMBER: US/09/220,920
; CURRENT FILING DATE: 1998-12-24
; EARLIER APPLICATION NUMBER: 09/163,283
; EARLIER FILING DATE: 1998-09-29
; EARLIER APPLICATION NUMBER: 60/108,148
; EARLIER FILING DATE: 1998-11-12
; EARLIER APPLICATION NUMBER: 09/218,698
; EARLIER FILING DATE: 1998-12-22
; NUMBER OF SEQ ID NOS: 120
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 116
; LENGTH: 201
; TYPE: PRT
; ORGANISM: Homo sapiens

US-09-220-920-116
Alignment Scores:
Pred. No.: 3.31 Length: 201
Score: 87.50 Matches: 55
Percent Similarity: 31.75% Conservative: 12
Best Local Similarity: 26.07% Mismatches: 93
Query Match: 7.80% Indels: 51
Gaps: 8
DB: 10


```

Db      24  AlAGlyClySerSer-----SerGlyThrValAsnProValProThrThrAlaSerPro 41
QY      521  GGGCGGCT-----CGGCAACTGGCGCTCCCGCTCAAAATCGATGGAATGAGCGCTCA 468
Db      42  IleSerSerIlePheSerArgThrGlyAspProAlaProAsnMetGluAsnThrThrSer 61
QY      467  GGC-----465
Db      62  GlyPheLeuGlyProLeuLeuValLeuGlnAlaGlyPhePheLeuLeuThrArgIleLeu 81
QY      464  -----TGGGGTTGGGTTAATCAAAATCAGTTGGGTGGC---432
Db      82  ThrIleProGlnSerLeuHisSerTrp---TrpThrSerLeuAsnPheLeuGlyAlaAla 100
QY      431  ---CAGTTGTGGNAAGTCAAAATTCGTGCGGGTGGACTCCACCACTCCGGGGCGGTGC 375
Db      101  ProThrCysLeuGlyGlnAsnSerGlnSerProThrSerAsnHisSerProThrSerCys 120
      374  TGGATCACATCTCTGTTGACGCCCATCTCTTCAGCGCTCTCTCCAAAGGTACTGCTTGATC 315
      121  -----ProIleCys-----124
QY      314  TTGTCAAAGTGGCGGTGGTCCAGGCGCTGGCCCTTCACCAGATGGCGGTGTCGCTCGTAC 255
Db      125  -----ProGlyTyrArgTrpMetCysLeuArgArg 134
QY      254  ATGCTTCGGCCCTTGATTGGTCTGCTCGGCCAAACACGCAAGCTCATGAACCTTGACCTGC 195
Db      135  PheIleIlePheLeuPhe-----IleLeuLeuLeuCys 145
QY      194  TTCATCTTC-----TGCTCTTGCATG 174
Db      146  LeuIlePheLeuLeuValLeuLeuAspTyrGlnGlyMetLeuSerValCysProLeuLeu 165
QY      173  TCCAGGACTCGAAGAGGCGACAGCTCCGGGTGAGCCAGCACCTTATCGTAGAAGGTG 114
Db      166  ProArgThrSerThr-----ThrSerThrGlyProCysLysThrCysThrIleProAla 183
QY      113  TCAACCGCCAGCTTCATGCTTCTGCG---CGGCCAGGTCATCAACAGCTTCTTGGCG 57
Db      184  GlnGlyThrSerMetPheProSerSerCysCysThrLysProSerAsp-GlyAsnCysTh 203
QY      56  CCGCATCCGCTTCGCCGCTCT-----CTGTGCTGCTGG 23
Db      203  rCysIleProIleProSerSerTrpAlaPheAlaArgPheLeuTrpGluTrp 220

```

arch completed: April 16, 2003, 13:03:25
 job time : 29 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 23, 2003, 11:24:35 ; Search time 2154 Seconds
(without alignments)
8079.609 Million cell updates/sec

Title: US-09-920-953-2
Perfect score: 598
Sequence: 1 ggcgtcctgcagacgcggc.....agccatttcgaccacgaagcc 598

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 1451402878 residues
Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl:

- 1: gb_ba:*
- 2: gb_htg:*
- 3: gb_in:*
- 4: gb_om:*
- 5: gb_ov:*
- 6: gb_pat:*
- 7: gb_ph:*
- 8: gb_pl:*
- 9: gb_pr:*
- 10: gb_ro:*
- 11: gb_sts:*
- 12: gb_sy:*
- 13: gb_un:*
- 14: gb_vi:*
- 15: em_ba:*
- 16: em_fun:*
- 17: em_hum:*
- 18: em_in:*
- 19: em_mu:*
- 20: em_om:*
- 21: em_or:*
- 22: em_ov:*
- 23: em_pat:*
- 24: em_ph:*
- 25: em_pl:*
- 26: em_ro:*
- 27: em_sts:*
- 28: em_un:*
- 29: em_vi:*
- 30: em_htg_hum:*
- 31: em_htg_inv:*
- 32: em_htg_other:*
- 33: em_htg_mus:*
- 34: em_htg_pln:*
- 35: em_htg_rtd:*
- 36: em_htg_mam:*
- 37: em_htg_vrt:*
- 38: em_sy:*
- 39: em_htgo_hum:*
- 40: em_htgo_mus:*
- 41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	81.4	13.6	1372	8	CELI410	X72915 C.eugametos
2	79.7	13.2	587	3	TETHEMOGP	D13920 Tetrahymena
3	77.6	13.0	1197	8	CELI637P	X72916 C.eugametos
4	76.6	12.8	481	1	AF475938	AF475938 Synecocyst
5	70.4	11.8	146174	1	D90910	D90910 Synecocyst
6	53.8	9.0	494	3	TETHEMOGT	D13919 Tetrahymena
7	50.8	8.5	3605	1	MKNATNS	Z50745 M.kandleri
8	50.8	8.5	12792	1	AE010389	AE010389 Methanopy
9	48.2	8.1	1898	10	MUSDELTA	M74590 Mouse delta
10	48.2	8.1	2330	10	MUSUCRBP	L13969 Mouse delta
11	48.2	8.1	3041	10	MUSSTRANS01	AL653048 Mouse delta
12	48.2	8.0	234817	10	LMFLCHR32_06	Continuation (7 of 8)
13	47.6	8.0	110000	2	LMFLCHR36_07	Continuation (8 of 8)
14	47.6	7.9	137936	2	CNS08CAX	AL844874 Oryza sat
15	47.4	7.9	147778	2	CNS07EFR	AL513404 Oryza sat
16	47.4	7.9	3014	3	AY084205	AY084205 Drosophill
17	47.2	7.9	10486	1	BJDNAKJ	Y09633 Bradyrhizob
18	47.2	7.8	411	1	MSM249386	AJ249386 Mycobacte
19	46.4	7.8	1737	1	AF130980	AF130980 Mycobacte
20	46.4	7.8	16563	1	AE007025	AE007025 Mycobacte
21	46.4	7.8	35377	1	MTCY48	274020 Mycobacteri
22	46.4	7.8	117080	2	AC098840	AC098840 Magnaport
23	46.4	7.7	135216	8	AC069145	AC069145 Oryza sat
24	45.8	7.7	144724	8	AC078948	AC078948 Oryza sat
25	45.8	7.6	149117	2	AC120527	AC120527 Oryza sat
26	45.2	7.5	11401	1	D50308	D50308 Lysobacter
27	45	7.5	5520	3	AC084329	AC084329 Leishmani
28	44.8	7.5	11034	1	AE005065	AE005065 Halobacte
29	44.6	7.4	154494	2	AP005070	AP005070 Oryza sat
30	44.4	7.4	11509	1	AE012381	AE012381 Xanthomon
31	44.2	7.4	12172	1	AE012305	AE012305 Xanthomon
32	44	7.4	79390	2	AC020442	AC020442 Drosophill
33	44	7.4	84653	3	AC004290	AC004290 Drosophill
34	44	7.4	167350	2	AC129717	AC129717 Oryza sat
35	44	7.4	180814	3	AC099023	AC099023 Drosophill
36	44	7.4	256764	3	AE003792	AE003792 Streptomy
37	44	7.3	32274	1	SCC8A	AF145230 Ralstonia
38	43.8	7.3	867	1	AF145230	Continuation (4 of 5)
39	43.4	7.3	110000	2	LMFLCHR36_03	AE001826 Deinococ
40	43.4	7.3	177466	1	AE001826	AF084543 Tupaia m
41	43.4	7.2	13461	14	AF084543	AF281817 Tupaia m
42	43.2	7.2	195859	14	AF281817	AC124194 Mus muscu
43	43.2	7.2	201271	2	AC124194	AL646067 Ralstonia
44	43	7.2	213050	1	AL646067	
45	43					

ALIGNMENTS

RESULT 1	CELI410	1372 bp	mrna	linear	PLN 16-JUN-1994
LOCUS	C.eugametos				
DEFINITION	X72915				
ACCESSION	X72915				
VERSION	X72915.1				
KEYWORDS	haemoglobin; LI410 gene; light-induced expression; nuclear gene.				
SOURCE	Chlamydomonas eugametos.				
ORGANISM	Chlamydomonas eugametos				
REFERENCE	Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;				
AUTHORS	Chlamydomonadaceae; Chlamydomonas.				
	1 (bases 1 to 1372)				
	Couture,M., Chamberland,H., St-Pierre,B., Lafontaine,J. and				
	Guertin,M.				
TITLE	Nuclear genes encoding chloroplast hemoglobins in the unicellular				

green alga Chlamydomonas eugametos
 JOURNAL Mol. Gen. Genet. 243 (2), 185-197 (1994)
 MEDLINE 94322186
 PUBMED 8177215
 REFERENCE 2 (bases: 1 to 1372)
 AUTHORS Guertin, M.
 TITLE Direct Submission
 JOURNAL Submitted (23-MAR-1993) M. Guertin, Laval University, Department of
 Biochemistry, Pavillon Vachon Room 3426, Quebec, G1K 7P4, CANADA
 FEATURES Location/Qualifiers
 source
 1..1372
 /organism="Chlamydomonas eugametos"
 /strain="DTEX9"
 /db_xref="taxon:3053"
 /clone="lambda 21.1.1"
 /clone_lib="cDNA library in lambda ZAP"
 84..599
 /gene="LI410"
 84..599
 /gene="LI410"
 /codon_start=1
 /product="Li410p"
 /protein_id="CA51420.1"
 /db_xref="GI:437981"
 /db_xref="SWISS-PROT:P52334"
 /translation="MKRTVQLTLPICRAQQOVPVRAATATTPATKKCPFS
 LFAKLGGEAVEADVDFYKRVADPTVSKDMVQRSKQFAFAYALGGAWE
 KGRMRTAKHDLVPHLTDVHFQAVVYHLSDTLAEILGVTPGDIADAMAVVASTKTEVLN
 MKPQQGAESEN"
 BASE COUNT 278 a 421 c 377 g 296 t
 ORIGIN

Query Match 13.6%; Score 81.4; DB 8; Length 1372;
 Best Local Similarity 52.1%; Pred. No. 2.5e-06;
 Matches 209; Conservative 0; Mismatches 186; Indels 6; Gaps 1;
 QY 1 GCCTGCTGCACAGCGGCCACAGCAGAGCGGAGCGGATCGGGCGC 60
 DB 150 GCCTGCTGCACAGCGGCCACAGCAGAGCGGAGCGGATCGGGCGC 209
 QY 61 AAGAAGCTGTTGATGACCTGGCGCGCGCAGAGGATGAAGTGGCGTGGACATTC 120
 DB 210 TTCTCACTCTTCGCCAAGCTGGCGCGCGGAGCGGATGAAGTGGACATTC 269
 QY 121 TACGTAAGTGTGCTGACCGCGGAGCTGCTCCCTTCTTCAGTCCCTGGACATGCAA 180
 DB 270 TACAACAAGTGTGCGCGACCCCGCTCAGCGTCTTCTTCCAGAGCGGACATGAAG 329
 DB 181 GAGCAGAAGATGAAGAGGTCAAGTTTCATGAGCTTCTGTTGGCGGAGCAGACCAATAC 240
 DB 330 GTCCAGCGCTCAAGCAGATTTGCTGCTGCGCTATGCTCTGGTGGTGGCAGCGAGTGG 389
 QY 241 AAGGCGCGAAGCATGTACAGCAGCAGCGCCATCTGTCAAGGCGCCAGCGCTGGACAC 300
 DB 390 AAGGCGAAGCATGTGCGACCGCAGCAGCAGGACCTGCT-----CCCTCAGCTCAGCAG 443
 QY 301 CGGCATTTGACAAGATCAAGCAGTACCTTGGAGAGACGCTGCAAGAGATGGCGCTCAG 360
 DB 444 GTCCACTTCCAGCGCGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 503
 QY 361 CAGGATGTGATCAGCAGCGCGCGGAGTGTGGAGTCCAC 401
 DB 504 CCTGGGACATCGCTGATGCCATGGCAGTGGTGGCTCTAC 544

RESULT 2
 TETHEMOGP
 LOCUS TETHEMOGP 587 bp mRNA linear INV 27-OCT-2000
 DEFINITION Tetrahymena pyriformis mRNA for hemoglobin.
 ACCESSION D13920
 VERSION D13920.1 GI:217409
 KEYWORDS hemoglobin.
 SOURCE Tetrahymena pyriformis cDNA to mRNA.

ORGANISM

Tetrahymena pyriformis
 Eukaryota; Alveolata; Ciliophora; Oligohymenophorea;
 Hymenostomatida; Tetrahymenina; Tetrahymena.
 REFERENCE 1 (bases 1 to 587)
 AUTHORS Takagi, T., Iwasa, H., Yuasa, H., Shikama, K., Takemasa, T. and
 Watanabe, Y.
 TITLE Primary structure of Tetrahymena hemoglobins
 JOURNAL Biochim. Biophys. Acta 1173 (1), 75-78 (1993)
 MEDLINE 93250050
 REFERENCE 2 (bases 1 to 587)
 AUTHORS Takagi, T.
 TITLE Direct Submission
 JOURNAL Submitted (09-DEC-1992) Takashi Takagi, Biological Institute,
 Faculty of Science, Tohoku University; Aobayama, Sendai, Miyagi
 980, Japan (E-mail: TAKAGI@PNTOHOK.BITNET,
 Tel:022-222-1800(ex.3469), Fax:022-263-9206)
 FEATURES Location/Qualifiers
 source
 1..587
 /organism="Tetrahymena pyriformis"
 /db_xref="taxon:5908"
 17..382
 /codon_start=1
 /transl_table=6
 /product="hemoglobin"
 /protein_id="BAA03015.1"
 /db_xref="GI:418106"
 /translation="MNKPQTIYEKLGGENAKAAVPLFYKVKVLADERVKHFKNFTDMD
 HOTQQTDFLMLLGGPNHYKKNMTEAHKGNLQNLHFDALIEENLAATLKLGLVTD
 VINEAKVIEHTRKDLGK"
 polyA_signal 442..447
 polyA_signal 482..487
 BASE COUNT 198 a 121 c 104 g 164 t
 ORIGIN

Query Match 13.2%; Score 79; DB 3; Length 587;
 Best Local Similarity 54.1%; Pred. No. 8.2e-06;
 Matches 191; Conservative 0; Mismatches 150; Indels 12; Gaps 1;
 QY 58 CGCAGAAGCTGTTGATGACCTGGCGCGGAGGATGAAGTGGCGGTGGACAC 117
 DB 26 CCCCAACTATTATTGAAAAGCTCGGAGCGGAAAATGCCATGAAGCTGCCCTCTC 85
 QY 118 TTCTAGCATAAGTGTGCTGCTGACCGGAGCTGCTCCCTTCTTCAGTCCCTGGACATG 177
 DB 86 TTCTACAAGAAGTCTTAGCTGATGAAGAGTCAAGCATTTCTTCAAGAACACCGACATG 145
 QY 178 CAAGACAGAAGATGAAGCAGGTCAAGTTCATGAGCTTCTGTTGGCGGAGCAGACCAA 237
 DB 146 GATCACCACCAAGCAATAAAGTACTTCTCACCACATGCTTAGTGGTCCCAACCAT 205
 QY 238 TACAGGGCGGAGCATGTACGACGACAGCGCCCATCTGGTCAAGGGCCACGGCCCTGGAC 297
 DB 206 TACAAGGGTAAATAATATGACTGAAGCTCA-----CAAGGGTATGAATTCGAA 253
 QY 298 CACCGGCCTTTGACAAGATCAAGCAGTACCTTGGAGAGACGCTGCAAGAGATGGCGCTC 357
 DB 254 AACTTGCATTTGATGCCCATATTGAAACCTTGTGCTACCTTAAAGAGCTCGGTGTC 313
 QY 358 AAGCAGGATGTGATCCAGCAGCGCGCGGAGTGGTGGAGTCCACCGCCGACGA 410
 DB 314 ACCGATGCTGTTATTAAACGAGGCTGCTAAGTCAATCAACACACCGCTAAGGA 366

RESULT 3
 CELI637P
 LOCUS CELI637P 1197 bp mRNA linear PLN 16-JUN-1994
 DEFINITION C.eugametos LI637p mRNA.
 ACCESSION X72916
 VERSION X72916.1 GI:437982
 KEYWORDS haemoglobin; LI637 gene; light-induced expression; nuclear gene.
 SOURCE Chlamydomonas eugametos.
 ORGANISM Chlamydomonas eugametos
 Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;